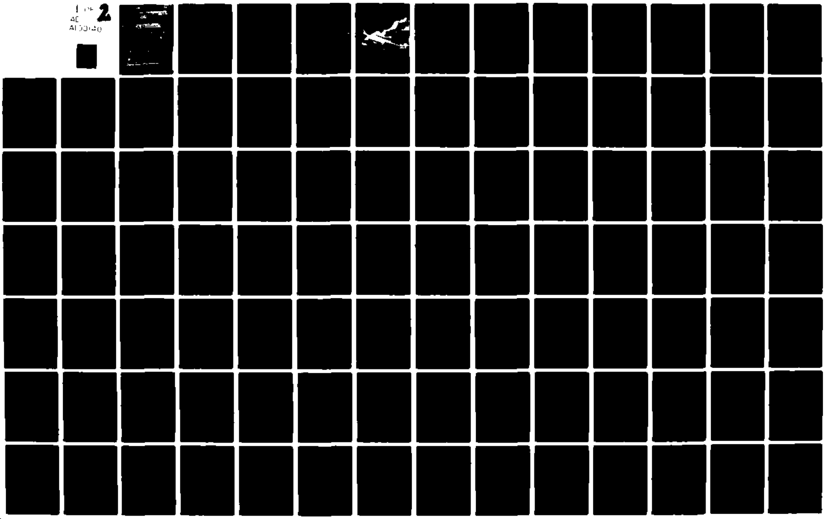
AD-A100 140 PERCEPTRONICS INC WOODLAND HILLS CALIF F/G 5/9
AIRCREW EMERGENCY DECISIONS: ANALYSIS OF CUE FACTORS FOR USE IN--ETC(U)
DEC 80 L F LUCACCINI, R HOPF-WEICHEL F49620-78-C-0067
UNCLASSIFIED PFTR-1065-80-12 AFOSR-TR-81-0475 NL

1 OF 2
AD
A100140



AD A100140

LEVEL

AIRCREW EMERGENCY [illegible]
ANALYSIS OF CUE FACTORS [illegible]
IN SYNTHETIC LEARNING [illegible]

Luigi F. Lucaccini
Rosemarie Hopf-Weichel
Francois G. Christen
Amos Freedy

DTIC
ELECTE
JUN 1 2 1981
C

Prepared for:
AIR FORCE OFFICE of SCIENTIFIC RESEARCH
Building 410
Bolling AFB, D.C. 20332

FILE COPY

PERCEPTRONICS
[illegible] VARIEL AVENUE • WOODLAND HILLS • CALIFORNIA [illegible] • PHONE [illegible]

Approved for public release;
distribution unlimited.

81 6 12 086

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 81-0475 — 2. GOVT ACCESSION NO.: AD-A100 140	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Aircrew Emergency Decisions: Analysis of Cue Factors for Use in Synthetic Learning Environments	5. TYPE OF REPORT & PERIOD COVERED: Final
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Luigi Lucaccini, Rosemarie Hopf-Weichel, Francois Christen, and Amos Freedy	8. CONTRACT OR GRANT NUMBER(s): F49620-78-C-0067
9. PERFORMING ORGANIZATION NAME AND ADDRESS: PERCEPTRONICS, INC. 6271 Variel Avenue Woodland Hills, CA 91367	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2313/A2
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research /NL Building 410 Bolling AFB, D.C. 20332	12. REPORT DATE: December 31, 1980
	13. NUMBER OF PAGES: 176
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Aircraft Emergencies	Cue Characteristics	Decision Processes
Aircraft Malfunctions	Cue Patterns	Decision Training
Boldface	Decision Making	Evaluation
Cognitive Factors	Decision Models	Prototype Instruction
Cues	Decision Procedures	Situational Emergency Training

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes the second year activities of a program of research and development to ennance pilot and aircrew emergency decision skills. Major objectives of the work carried out were to: (a) extend theoretical models developed earlier, (b) to carry out four experimental studies which would yield basic pilot judgment data for cognitive and situational factors involved in cue patterns typical of aircraft emergencies, (c) to design a scenario generation system utilizing cognitive and situational factors, and (d) to present a general approach for training and assessing aircrew emergency decision skills. The

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. (Continued)

work of the second year of this program focused heavily on the specification of relationships of cues and cue patterns as decision process inputs in the complex, high performance aircraft setting.

Studies 1 and 2 assessed the feasibility of using specified cue characteristics as variables impacting expert pilot judgments and decisions. In addition, the nature of interactions among cues presented in patterns and the impact of cue patterns were experimentally determined. A number of interesting results were obtained which facilitate control of the inter-cue interaction and its effects on expert judgment. Situational cues were found to operate in the same fashion as cues related to specific aircraft malfunctions, allowing the incorporation of both types of decision information in a common framework.

Study 3 was concerned with validating predictions of cue impact on decision making. Difficulty of decision making was shown to be related to cue characteristics. Similarly, the time and safety criticality features of the cue configuration were found to affect the breadth and amount of decision activity. A follow-up experiment, Study 4, considered the impact of additional cue patterns on pilot judgments. Differences between experienced and relatively inexperienced pilots in terms of judgments related to decision making were also investigated.

Meta-evaluation theory and principles of instructional system evaluation are examined in the context of decision training in order to identify theoretical and practical considerations for evaluating aircrew decision training. A prototype instructional approach is presented in the form of an emergency decision situation which embeds elements of a decision process model.

12

Final Technical Report PFTR-1065-80-12
Contract No. F49620-78-C-0067
Dec[redacted]980

AIRCREW EMERGENCY DECISIONS: ANALYSIS OF CUE FACTORS FOR USE IN SYNTHETIC LEARNING ENVIRONMENTS

DTIC
ELECTE
JUN 1 2 1981
S D
C

Luigi F. Lucaccini
Rosemarie Hopf-Weichel
Francois G. Christen
Amos Freedy

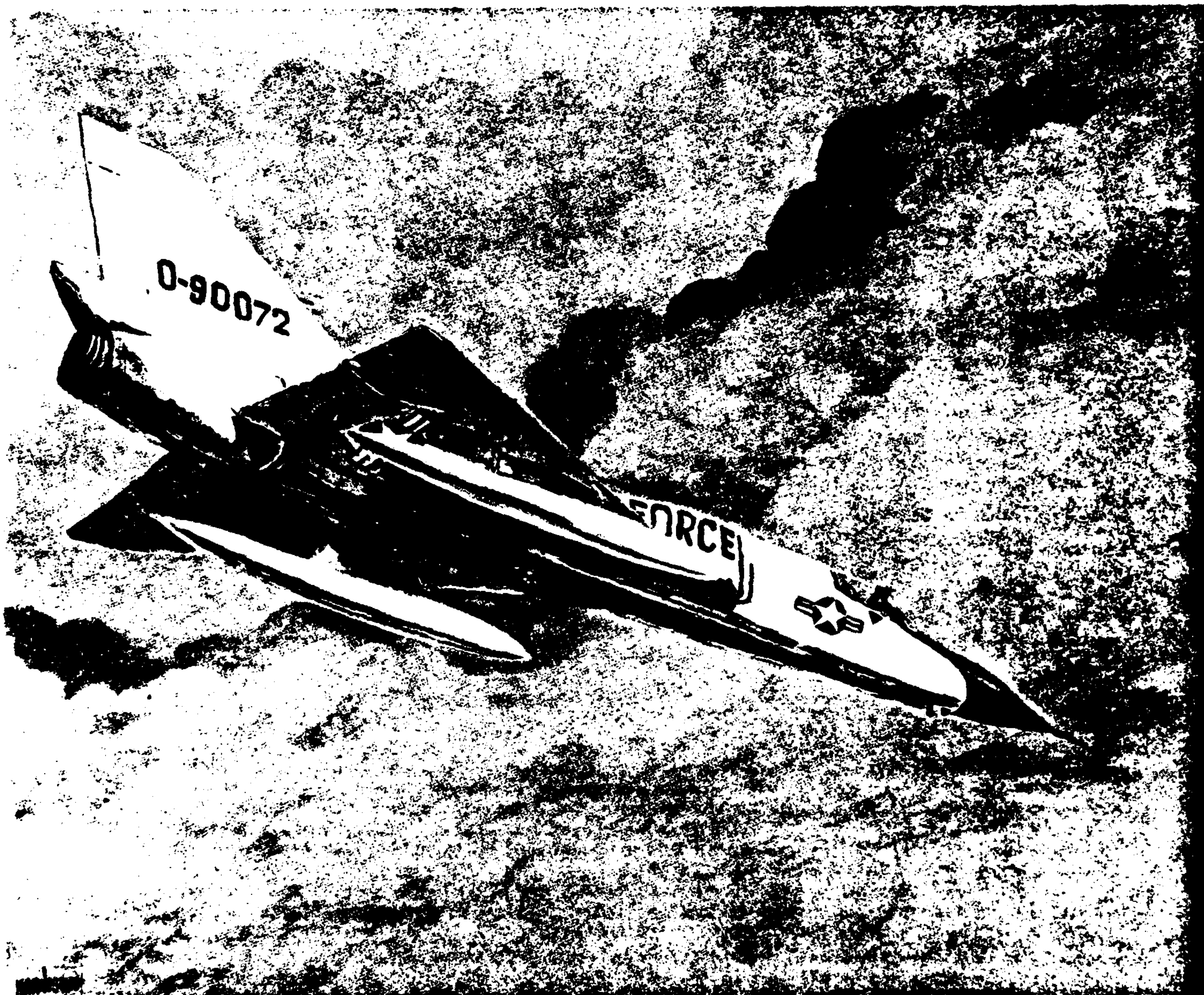
Prepared for:

AIR FORCE OFFICE of SCIENTIFIC RESEARCH
Building 410
Bolling AFB, D.C. 20332

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b)
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

PERCEPTRONICS

6271 VARIEL AVENUE • WOODLAND HILLS • CALIFORNIA 91367 • PHONE (213) 884-7470



EMERGENCY AIRCREW DECISIONS

TABLE OF CONTENTS

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Avail and/or
Dist Special
A

TABLE OF CONTENTS (CONT.)

TABLE OF CONTENTS (CONT.)

1. OVERVIEW

1.1 Objectives of Year 2

This report describes the activities and findings of the second year of a research and development program to enhance pilot and aircrew emergency decision skills. The overall effort is intended to provide a basis for designing training and assessment techniques which address the development of aircrew emergency decision skills in a more systematic and comprehensive manner than has been possible in the past.

Major objectives of Year 2 were:

(1) To revise and validate the models developed in Year 1 through collection of data for a specific aircraft system.

(2) To carry out experimental studies which would yield basic pilot judgment data for cognitive and situational factors involved in the cues and cue patterns typical of aircraft emergency decisions.

(3) To design a scenario generation system which allows the construction of aircraft emergency scenarios utilizing cognitive and situational factors.

(4) To specify a general approach for the training and assessment of aircrew emergency decision skills.

1.2 Theoretical Background and Problem Statement

Perceptronics' basic research program on Situation Emergency Training (SET) focuses on the acquisition of complex judgment and decision skills in the high performance context of the aircraft emergency.

An aircraft emergency is commonly defined as the unexpected occurrence of a set of circumstances which calls for immediate judgment and action to avoid undesirable consequences. The standard emergency response expected of every aircrew is three-fold: (a) to maintain aircraft control, (b) to analyze the situation and take proper action, and (c) to land as soon as practicable. In the broader context of flight safety, aircrews are expected to do more than skillfully resolve immediate full-blown emergencies. It is equally important that they actively avoid situations which might lead to emergencies and that they recognize the early signs of an impending emergency and take corrective action before the situation assumes crisis proportions. An aircraft emergency might be viewed, then, as a sequence of events and decisions, which, if not appropriately resolved at an earlier stage, culminates in a crisis. If so, the study of aircraft emergencies and the design of emergency decision training must accommodate the broader range of situations and skills that this conception embraces.

In one sense, decision making can be viewed as the efficient translation of information into appropriate action by a rational decision maker using effective decision strategies; however, this may be a better description of decision analysis. A slightly broader view of decision making is necessary to encompass the decision activities important in the complex performance setting of aircrew emergency decision problems. Clearly, decision making in this case involves a number of overlapping aspects or phases of activity, which might be conceived of as a series of related problem solving tasks. In our view, the following general breakdown is useful in characterizing decision making in the complex (emergency) setting:

(1) Situation Diagnosis.

(a) Problem recognition.
(b) Information acquisition and evaluation.

(2) Decision Making.

  (a) Problem structuring and development of alternatives.
  (b) Evaluation of alternatives and selection of a course of action.

(3) Decision Execution.

  (a) Implementation of action alternative.
  (b) Monitoring of implementation and evaluation of results.

The impact of situational factors on emergencies and emergency decisions can not be ignored. A decision which is appropriate under one set of circumstances can be fatal under another. Aircrew members must be prepared not only to exercise the wide range of decision skills suggested above, but to consider relevant situational factors appropriately in formulating and executing emergency decisions.

Aircraft emergencies clearly represent an important performance problem area worthy of study in and of itself because of the far-reaching practical implications of aircraft emergencies and accidents. What is sometimes overlooked, however, is the opportunity that this complex and highly demanding performance setting offers for the study and resolution of basic theoretical issues in human judgment and decision making. Some of these basic research questions include:

(1) What models of judgment and cognitive processes most accurately and completely describe decision making in complex, high performance settings?
(2) What theoretical models can be developed to specify the relationship of event-related and contextual cues to decision making and the efficiency and quality of decision?

(3) How can the decision skills and strategies of novices and experts best be observed, measured, and evaluated in the complex, high performance setting?

(4) What theoretical guidance can be derived for the development of learning principles to facilitate acquisition of complex, high performance judgment and decision skills?

(5) How can synthetic learning environments best be assembled to enhance development of decision skills appropriate to the complex, high performance setting?

These basic research questions are particularly challenging in view of our currently limited understanding of human judgment and decision making in general. Complex decision performance, such as is involved in emergency settings, is even less well understood and has had very little study from the standpoint of the behavioral decision sciences to date. In addition, little is known of how experts view decision tasks and how to elicit the knowledge, rules and judgmental processes experts use, particularly in complex settings. Finally, the emergency decision setting involves a class of decision problems and associated cognitive tasks which are often performed under time constraints, high stress, and with limited decision information and resources. Theoretical advances made in the study of decision activity and decision making in the aircraft emergency setting will be applicable, therefore, to the understanding of basic behavioral decision processes in other domains which involve similar complex, high performance requirements.

1.3 Accomplishments

1.3.1 Year 1: Activities and Accomplishments. A number of activities were carried out in the first year of this program in order to develop the theoretical and empirical knowledge base needed to support the succeeding years' efforts. These activities included reviews of SET and Boldface approaches to emergency decision training; site visits to a number of squadrons for observation, orientation, and interviews with flying pesonnel; reviews of flight manuals and other written materials for selected military aircraft; an extensive program of field data collection and the analysis of judgmental data derived from studies using expert flying personnel; a review and analysis of USAF aircraft accident reports for 1977; and the convening of a workshop of civilian and military personnel to review approaches to the analysis of aircraft emergency decisions, aircraft accident research, and aircrew emergency decision training.

Two major reports were prepared as a result of these activities. The first, Aircrew Emergency Decision Training: A Conference Report, summarizes the results of the workshop held in San Francisco in November 1978. The second is the Technical Report (PATR-1065-79-7) entitled Aircraft Emergency Decisions: Cognitive and Situational Variables, which summarizes the activities and findings of Year 1.

The work of Year 1 was successful in providing an initial theoretical basis for the analysis of aircrew emergency decision problems. A taxonomic structure was derived which appears to be of considerable value in specifying the cognitive aspects of aircraft emergency decision problems. There are two obvious applications of the taxonomy. The first is the specification of classes of emergency decision problems which are related in terms of the cognitive functions required for their successful resolution. The second area of application is at a more detailed level, namely in the development of scenarios or aircraft emergency decision problems for use in the training of pilots and other aircrew members.

The taxonomy incorporates both situational and task-specific elements as cognitive attributes of the decision tasks performed under emergency conditions. There are several steps preceding the development of the taxonomy. The aggregation of situations which could be considered within the context of complex human decision performance was reviewed. A definition of the emergency situation was developed which limited the scope of consideration to a manageable entity -- known malfunctions. Representational models of the objective (external) emergency situation, decision processes, and cognitive functioning were proposed as a way of characterizing the situational and behavioral aspects of an emergency malfunction. The taxonomic structure was then derived after consideration of the cognitive elements of the three representational models.

On the basis of the taxonomy, three classes of emergency situations were found to be of interest: Situation 1 (mostly predictable), Situation 2 (partly predictable), and Situation 3 (unpredictable). Initial guidelines for decision skill acquisition and evaluation are suggested in light of the cognitive requirements of each class.

The taxonomy also provides an initial framework for emergency decision problem generation. Situational and behavioral aspects of emergencies are covered at a level of detail which allows systematic identification of their cognitive elements. Thus, the utility of the taxonomy lies in the ability to correlate various and seemingly disparate elements of a given decision problem (or set of problems) in cognitive terms. This provides the capability to manipulate decision parameters in a systematic fashion so as to ensure that decision problems are described and specified in terms of a comprehensive and unifying factor -- cognitive functions.

Another key product of Year 1 is the preliminary classification of emergency situations according to the performance requirements of these situations as dictated largely by the physical nature of high-performance aircraft operations. This work facilitates the identification of emergency situations which are candidates for special emphasis in decision training programs. The classification, which derives from consideration of the risk, time pressures, and complexity of decision making tasks associated with specified malfunctions, lends an objective frame of reference to the theoretical tools provided by the representational models and the taxonomy.

1.3.2 Year 2: Activities and Accomplishments. A number of activities were carried out in Year 2 of this program to establish and extend a database of expert judgments derived from experimental studies and analyses. Major activities of Year 2 included revisions and expansions of theoretical models developed in Year 1 to describe the complex decision processes in the aircraft emergency performance setting; the development of initial conceptual linkages between event, decision and cognitive models for the complex, high performance setting; conducting a series of four experimental studies using a total of 50 F-4 expert (experienced) and novice (trainee) pilots to obtain basic judgment data specifying situational cue and cue pattern impact on decision performance; development of a formalized procedure for generation of synthetic learning environments for complex, high performance decision skill acquisition; initial specification of a training strategy and an initial test plan for subsequent testing of the effectiveness of these decision training concepts; and development of sample scenarios embodying these decision training concepts.

The work of Year 2 focuses heavily on the specification of relationships of cues and cue patterns as decision process inputs and decision information in the complex, high performance setting. The experimental validation through human judgment research of the decision process model and its linkages to situational cues should facilitate the design of synthetic learning environments which embody differing decision problems and decision performance demands. These theoretical achievements provide a basis for the specification of instructional strategies, instructional materials, and techniques to evaluate the decision skills and performance deficits of individuals carrying out judgmental tasks in high performance settings.

Studies 1 and 2 of Year 2 experimentally assessed the feasibility of using specified cue characteristics as variables impacting expert pilot judgments and decisions. In addition, the nature of interactions among cues presented in patterns and the impact of cue patterns were experimentally determined. A number of interesting results were obtained which facilitate control of the inter-cue interaction and its effects on expert judgment. Situational cues were found to operate in the same fashion as cues related to specific aircraft malfunctions, allowing the incorporation of both types of decision information in a common framework.

Study 3 of the Year 2 effort was concerned with validating predictions of cue impact on decision making. Difficulty of decision making was shown to be related to cue characteristics. Similarly, the time and safety criticality features of the cue configuration were found to affect the breadth and amount of decision activity. A follow-up experiment, Study 4, considered the impact of additional cue patterns on pilot judgments. Differences between experienced and relatively inexperienced pilots in terms of judgments related to decision making were also investigated.

1.4 Report Contents

Chapters 2 through 5 of this report describe in detail the results of the activities carried out in Year 2 of this effort. Chapter 2 contains a summary of the theoretical approach developed in Years 1 and 2 to the aircraft emergency decision situation as well as some initial training implications. Chapter 3 sets forth the findings from four studies conducted with beginning and advanced level F-4 pilots. The basic judgment data derived for six decision-related attributes of malfunction and situational cues are presented in some detail. Chapter 4 outlines a basic approach to evaluating prototype instructional materials, and considers the special requirements of training and evaluating decision skills. Chapter 5 presents a prototype instructional system for aircrew emergency decision training which derives from the decision process model developed earlier and which utilizes emergency scenarios composed of the types of malfunction-related and situational cues considered in Chapter 3.

2. THEORETICAL BACKGROUND

2.1 Introduction

As a background to the experimental work of Years 1 and 2, theoretical work was performed to establish a base for selection of variables and relationships for further study. This chapter summarizes the theoretical models and taxonomic structure developed and some related implications for training of aircrew emergency decision skills.

2.2 Representative Models

The following section describes three models which differentially relate aircrew emergency decision situations to behavioral variables. These three models are (1) an objective event model, corresponding to external events and representing an objective description of an emergency situation; (2) a decision process model describing the conscious processes needed to deal with an emergency, specifically the components of the decision-making situation; and (3) a cognitive process model, describing a theoretical view of the learning and memory processes that take place during training and in actual emergencies.

2.2.1 Event Sequence Model. Emergency situations can be the result of any number of factors, including those that are directly related to the pilot, such as physiological disturbances or psychological stress, communications break-downs, and so forth. Ideally, all types of emergencies should be included. For the time being, however, only malfunction-induced emergencies are considered. Similarly, causes or influences antecedent to a malfunction are excluded from the present conceptualization.

The event sequence model represents an objective view of the components which must be considered in developing training guidelines for emergency procedures. These are shown in Figure 2-1. By definition, the event sequence begins with a malfunction that is manifested by a pattern of cues. From the pilot's point of view, it is the information obtained from the cues that starts the sequence of dealing with an emergency. This information is defined by the values of the cue attributes. When cues are perceived and interpreted, they lead to certain actions that are described by the decision model and the cognitive model to be discussed below. These actions may lead to the identification of the malfunction or to intermediate outcomes that produce new cues until the problem is solved.

The cues must not only be interpreted for an immediate decision with respect to their cause and a possible solution to the emergency, but some anticipatory decisions must be made at this time, predicting the probable changes in the cues as a function of time and the consequences thereof. Both decision types are described below, and comprise the decision model.

2.2.2 <u>Decision Process Models</u>. For present purposes, two basic types of decisions will be distinguished. The first is the <u>ongoing decision</u> which requires immediate or continuing attention. The second is the <u>anticipatory decision</u> which may be executed at a later point in time. The two types of decisions are represented separately because they have some distinguishing characteristics, and becuase there seems to be an inherent difference between decisions concerning the problems brought about by a malfunction and decisions such as those involving ejections and abortive takeoffs. The first type--the ongoing decision--can be conceptualized as a classic decision, which includes problem structuring aspects as well as alternative selection and evaluation of outcomes. The second type involves anticipations concerning decisions that may have to be made at some future time, but because the time frame for

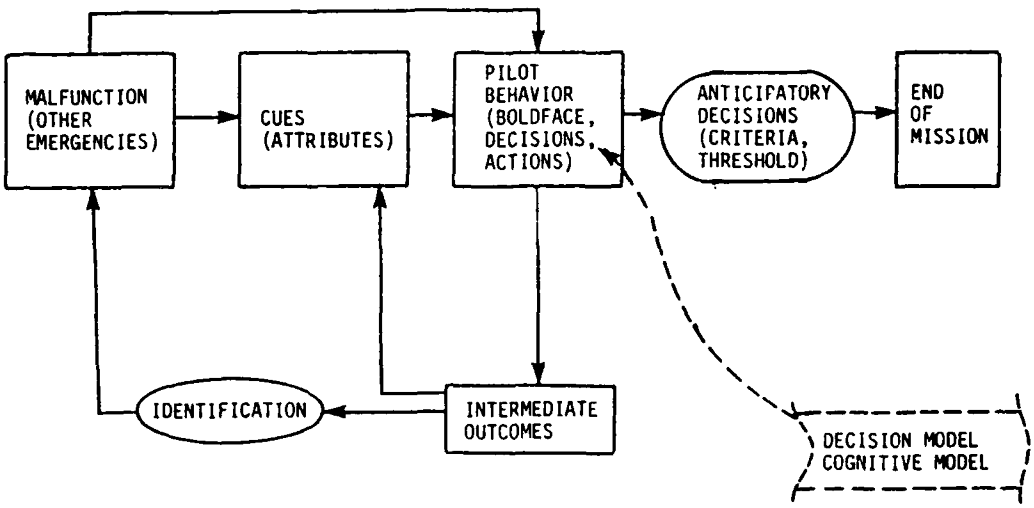


FIGURE 2-1.
EVENT SEQUENCE FOR EMERGENCY SITUATIONS

executing this type of decision is so critical, the conditions and criteria for executing it must be predetermined. The relevant components of these two types of decisions are shown in Figures 2-2 and 2-3, respectively.

Theoretically, the ongoing decision contains all the components identified in a complete decision (Figures 2-4 and 2-5), although in practice, some components may be irrelevent in specific cases. Two additional processes, hypothesis generation and confidence rating, are shown in Figure 2-2, because these may affect the manipulation and processing of the subsequent components. That is, it is assumed that problem recognition and structuring leads to the generation of hypotheses concerning the accuracy of this activity, and that the amount of confidence the decision maker has in how much of the problem has in fact been identified, is likely to affect the type and number of alternatives that will be examined, which, in turn, could influence subsequent phases of the decision process.

Figure 2-2 shows the stages of information processing of the ongoing decision. When a pattern of cues is perceived as deviating from the "normal" expected pattern of information, the deviation is a signal that something may be wrong. Problem recognition occurs, which is followed by problem structuring. The success of problem structuring is a function of the attribute values of the cues (reliability, salience, etc.), and these values in turn influence the degree of confidence a pilot has in his hypotheses concerning the malfunction. Very little is known concerning possible differences in how alternatives are formulated and selected, given different degrees of confidence in the problem structuring phase. Confidence is influenced by the degree of consistency among cues, and it determines the amount of additional information that is sought. The more inconsistency there is, the more information is needed to resolve it.

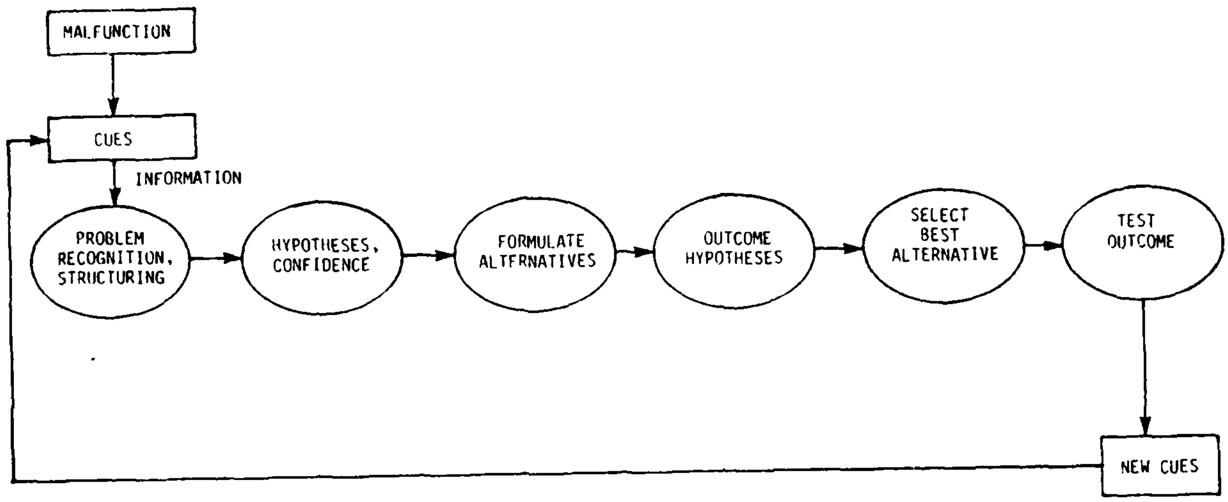


FIGURE 2-2.
COMPONENTS OF ONGOING DECISION

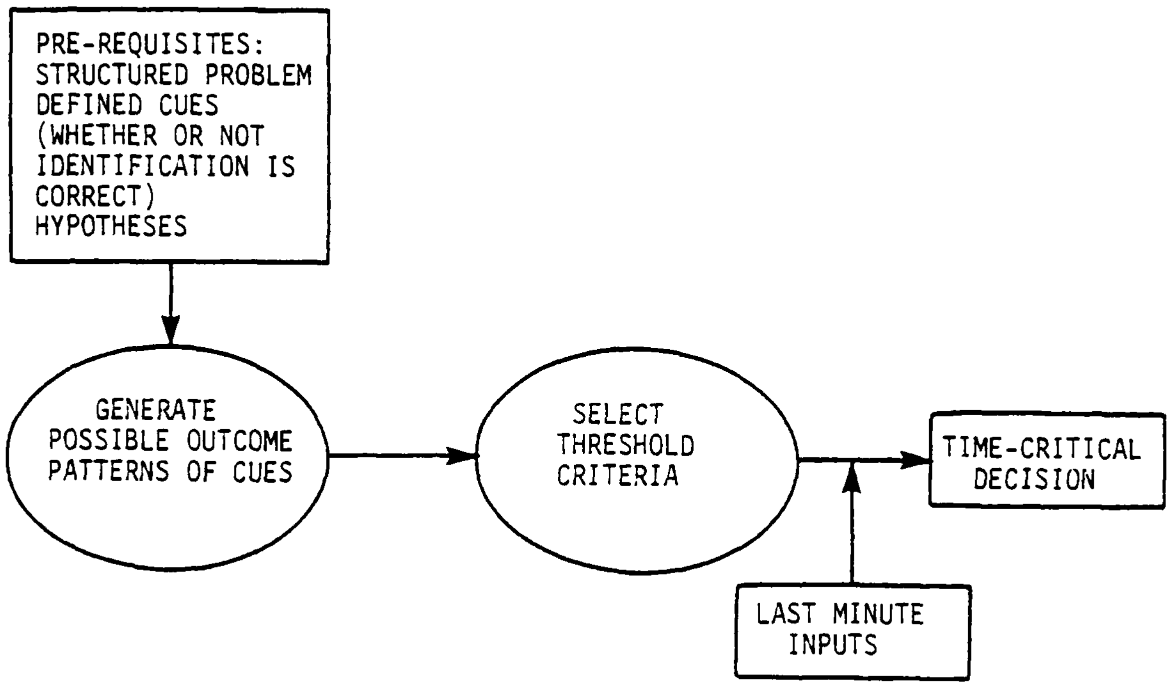


FIGURE 2-3.
COMPONENTS OF ANTICIPATORY DECISION

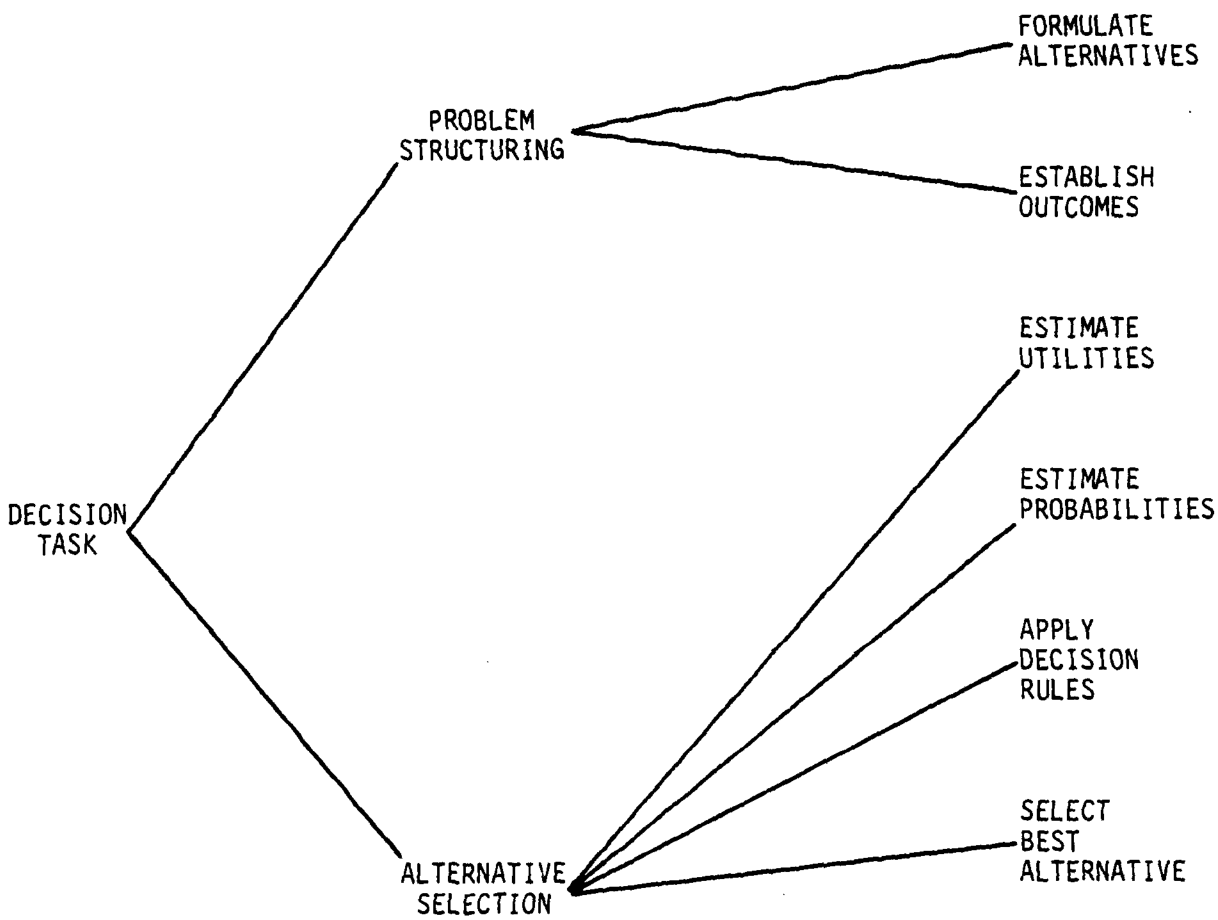


FIGURE 2-4.
DECISION TASK COMPONENTS

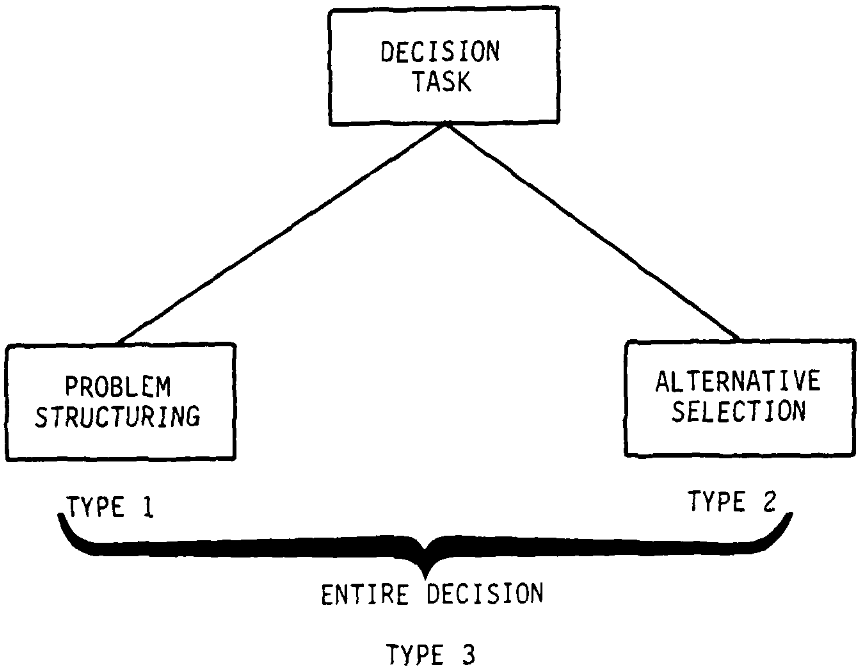


FIGURE 2-5.
DECISION TASK TYPES

When the alternatives have been formulated, hypotheses concerning the outcomes under each alternative are generated and a decision rule applied for selecting the best alternative. The outcomes following each action are tested against the hypotheses and the resultant information is used to start the decision loop anew until the problem is solved or the mission ended.

The anticipatory decision (Figure 2-3) begins when the cues have been identified and defined, and the problem has been structured. Confidence in the accuracy of this phase is not relevent at this point. What is relevant is that, given a malfunction and/or cues, there are certain probabilities associated with the way these cues can change over time and with the possibility of having to abort a mission, eject, or perform a forced landing. At the moment decisions such as these have to be executed, it is too late to perform a decision analysis. Therefore, the anticipatory decision is one in which the factors are anticipated and threshold criteria for executing the decision are preselected. Anticipatory decisions involve the generation of hypotheses concerning how the available cues may change over time and the specification of threshold values beyond which the changes require immediate action (e.g., ejection). Criteria for decision execution may also take new cues or changes in situational events into account.

The differences in the two classes of decisions are shown in Figure 2-6. For ongoing decisions dealing with malfunctions, the actions are discrete and determined by the cues as perceived at the moment. The outcomes are probabilistic in the sense that they may depend on factors that are not predictable, or they may depend on the accuracy of the problem recognition and structuring phase; the outcomes are also probabilistic because the estimated utilities determine the selection of the alternative actions. For anticipatory decisions, on the other hand, the actions are probabilistic

	ACTIONS	OUTCOMES	UTILITIES
ONGOING DECISIONS DEALING WITH MALFUNCTIONS	DISCRETE SET DETERMINED	PROBABILISTIC	ESTIMATED
ANTICIPATORY DECISIONS	THRESHOLD-DEPENDENT PROBABILISTIC	WELL-DEFINED	KNOWN

FIGURE 2-6.
DISTINCTIONS BETWEEN DECISION CLASSES

and the outcomes are determined. The actions are probabilistic because they depend on a critical threshold that may or may not be reached, and on the anticipated changes in the cues. Once the threshold is reached, the outcomes are well-defined since they are a function of the anticipatory decision. The outcome utilities do not affect the event alternatives since these are determined by extraneous factors that are not under the control of the pilot. The utilities are known, and the outcome is not a function of the pilot's decisions (although it is related to the type of malfunction that occurred and to how the pilot dealt with it).

The critical problem for anticipatory decisions is to recognize the changes in the relevent cues and the degree of changes that can be tolerated. Expert data obtained during interviews suggest that there is no objective way to define these changes, that they are a function of experience and "feel" for the aircraft. These same data also suggest that, while the threshold criteria may be successfully predetermined, the problem lies in the actual execution of the decision, especially in the case of a decision to eject. According to one expert pilot, there may be an interesting difference in this respect between experienced and inexperienced pilots. Although inexperienced pilots may know how to set the criteria, they may not follow through with their decisions or, for reasons such as lack of confidence and fear of repercussions, they may change their mind at the last minute. Experienced pilots tend to make the opposite error: once they make an anticipatory decision, they tend to stay with it, even when new information is obtained that would suggest a change. From a training point of view, it would seem desirable to investigate the possible reasons for this experience-related difference, and find out if it is possible to influence both tendencies--the one that delays the execution of the decision and the one that ignores new information.

The decision processes represented here can be trained directly, because these are processes of which the pilot is aware. Knowledge structures can be developed that include the relevant elements of decision making and the decision rules appropriate in specified circumstances.

2.2.3 Cognitive Process Model. The cognitive process model is shown in Figure 2-7. It is loosely patterned after the classic TOTE (test-operate-test-exit) unit of Miller, Galanter, and Pribram (1960) that assumes a feedback loop, whereby inputs from the environment are tested for congruity against some established criteria. If the test is positive, the input information is congruous with the information available in active memory, and an action can be taken. If the test leads to incongruity, additional information from long-term memory (LTM) must be activated until the results of the test are congruous.

In the present model, some additional assumptions are made concerning the organization of LTM and the processes whereby items are entered into active memory. LTM is assumed to consist of numerous systems, each representing a meaningful cluster of related information. These "representational systems" are analogous to the concept of "schema" (e.g., Bartlett, 1932; Hebb, 1949) or to Lashley's (1958) trace systems. All permanent events in memory belong to one system or another, but systems may overlap in varying degrees, share subsets of events, or be relatively autonomous. Also, systems vary on a number of dimensions, such as size (how much is known about a subject matter), stability (how reliably the knowledge can be retrieved), and complexity (how detailed the knowledge is). Both content-specific and procedural knowledge are included in a representational system, so that at this level no differentiation is made between structuring the problem in response to cues and dealing actively with an emergency.

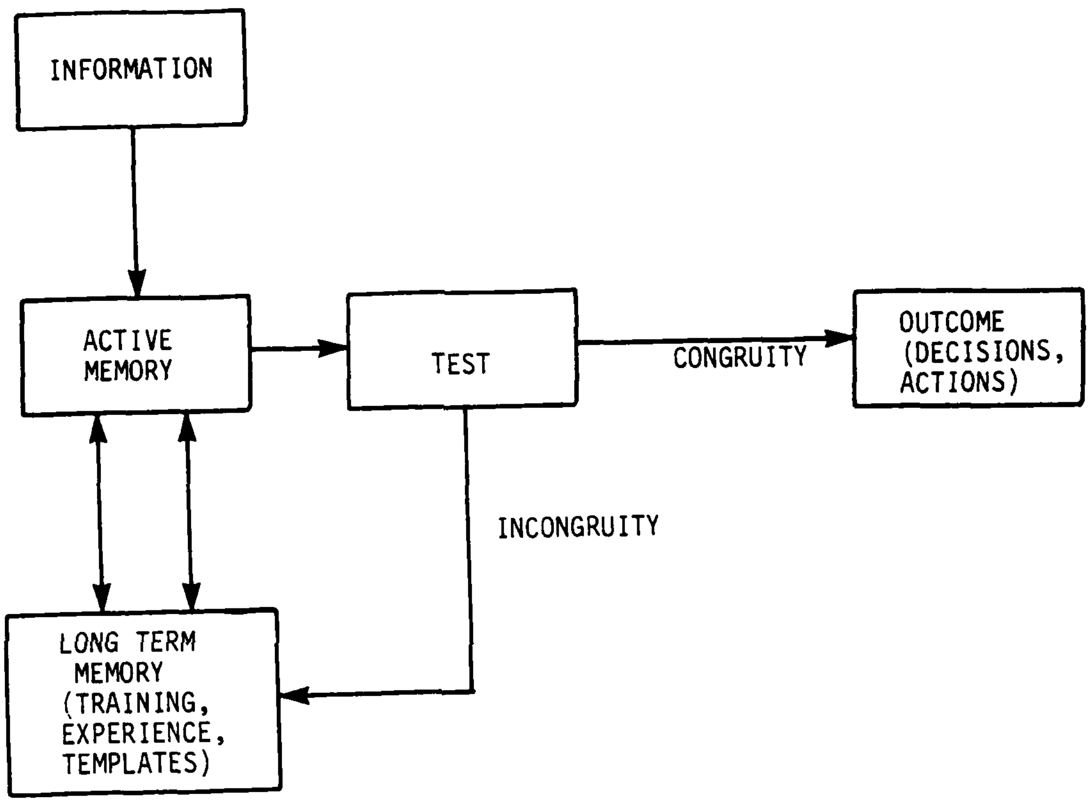


FIGURE 2-7.
COGNITIVE PROCESS MODEL

It is assumed that information in LTM is latent and that it must first be entered into active memory before it can be processed (Lashley, 1958). This implies that systems can only be altered when they are active; thus, learning (increasing the size of a system) and forgetting (decreasing the size of a system or its reliability) can only take place while the appropriate system is in active memory. It also implies that incoming stimuli (e.g., cues) can only be understood with respect to information belonging to systems that are active at the time. It is possible for several systems to be active at the same time, depending on their size and complexity, and a cue can be represented in more than one system at the same time; but in general, the most salient cue will activate the system that has its best representation. As a simplification, it is assumed that when relevant cues from the environment are perceived, they serve to activate the representational system that contains the information necessary to process the cue or to understand its meaning. The meaning of a cue is always understood with respect to the system that is active at the time, just as a homonym is interpreted with respect to its immediate context.

There are two ways that a system can be activated. The first was already mentioned, namely, through environmental stimuli that are perceived as being incongruous with those systems that are active at the time. This implies of course, that as long as one is conscious, there is always some system that is active; the problem as far as information processing is concerned, is how to switch from one system to another. The second way is internally. If an element in an active system also belongs to another (inactive) system, that element has the potential of activating the second system.

This conceptualization makes it possible to deal with the notion of "templates," their role in training and in dealing with emergencies, and their limitations. Templates may be defined as preplanned responses to emergencies. In the present formulation, a template is a special case of a representational system, which can only be activated when the pattern of the environmental stimuli matches all the elements in the system. It is extremely well-rehearsed and rigid in the sense that individual external elements are not likely to activate other systems. In other words, the correspondence between external events and the elements of the template is highly specific, and the system itself is relatively autonomous, so that it does not tend to activate other systems, and generally it can be activated only by a specified configuration of external stimuli. In this way, responses to these external stimuli are highly reliable and stress-resistant.

There will be times when templates become activated when in fact they do not represent the best solution to the situation. The problem for training is to teach pilots to recognize when the templates are applicable and how to activate the appropriate representational systems for dealing with unplanned emergencies when templates are not appropriate.

2.3 Basic Taxonomic Structure

2.3.1 Introduction. Four basic components have been identified as particularly important in describing an emergency situation and in developing training guidelines appropriate to those situations. The basic components include the situation as a whole, the malfunctions, the cues arising from the malfunction, and the behavior of the pilot. Each component is characterized by a set of attributes capable of assuming one of several values. These attributes identify the qualities of the components that are important in differentiating among emergency situations.

The structure of the taxonomy is not rigid. It is a preliminary attempt to categorize emergency situations in terms of the training needs that are anticipated. These needs will differ, depending on how specific and predictable the behavior of the pilot is, and whether the behavior includes psychomotor as well as procedural, cognitive, and decision-making behavior. As such, the taxonomy represents a general scheme for guiding training development.

2.3.2 Components and Their Attributes. The components of the taxonomy represent the major categories that have tc be described in terms of a set of attribute values. In the present context, attributes are not additive, rather they represent general features, used as differentiators of situational demands and to point to subsequent training requirements. The values assigned to the attributes may be either quantitative or qualitative, and serve to characterize different types of emergency situations.

The components and their attributes are shown in Figure 2-8. While there is an interrelationship between the attributes and their values across components, the values of each attribute are independently assigned to each component. For example, the description and values of the attribute "complexity" need not be the same when it pertains to malfunctions as when it pertains to cues. A brief description of the attributes follows. This is an initial attempt to describe relevant attributes, with alterations and additional refinements possible in particular applications.

2.3.2.1 Situation-Dependent Attributes. "Situation" refers to all external factors which affect the responses of the aircrew and the outcome of an emergency. This includes mission profile, flight phase, weather, time of day, communications, distance from help, and other similar factors.

(1) Predictability. All components have "predictability" as an attribute, and in each case, it will have the same description and the same values, namely, "mostly predictable," "not predictable," and "partially predictable." Predictability

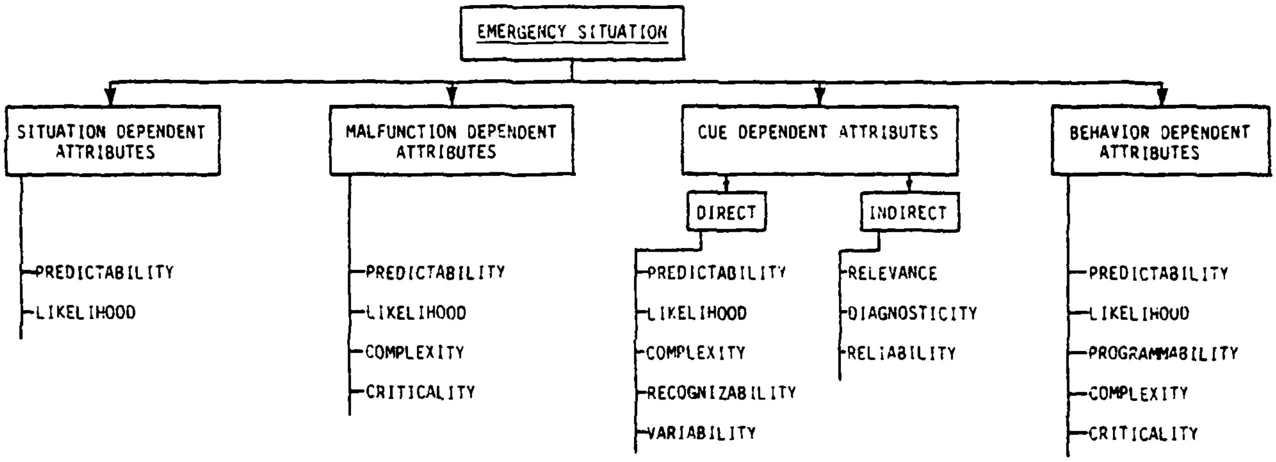


FIGURE 2-8.
COMPONENTS AND ATTRIBUTES OF EMERGENCY SITUATIONS

does not imply that the emergency itself can be predicted, but rather that it is possible to specify the influence of each situational factor on the outcome of the emergency.

(2) Likelihood. Likelihood refers to the probability that a particular situation will occur. This attribute is important in that it may suggest how much training time should be devoted to a certain combination of events. Likelihood is also an attribute that applies to all four components, and for which the attribute values will remain the same across components.

2.3.2.2 Malfunction-Dependent Attributes. Malfunctions refer to any physical breakdown, failure, or irregularity in the system. Attributes of malfunctions include:

(1) Predictability. See above.

(2) Likelihood. See above.

(3) Complexity. The complexity of a malfunction refers to the malfunction itself, its physical repercussions on other parts of the system, and the ease or difficulty with which it can be described. Single, compound, or sequential emergencies can exist.

(4) Criticality. This attribute refers to the potential criticality of the emergency situation in terms of its outcome. While both the situation and the cues can vary in criticality, the type of malfunction that exists is the major determinant of an emergency situation's outcome.

2.3.2.3 Cue-Dependent Attributes. A cue can be any information in the pilot's environment that is perceived through any of the senses. Because cues represent information concerning the malfunction, several of the attributes listed below refer to the relationship between the cues and the malfunction (indirect attributes), rather than to the cues themselves (direct attributes).

Direct Attributes

(1) Predictability.

(2) Likelihood. See above.

(3) Complexity. The complexity of cues is interpreted differently than the complexity of malfunctions. Malfunctions can be described objectively, whereas cues are only meaningful as they are perceived and interpreted by the pilot. Cues can be described in terms of the number used in diagnosing an emergency and their relative comparative strength or salience.

(4) Recognizability. Recognizability or salience refers to the probability that a particular cue will be perceived and to the time required to perceive it. Recognizability is a function of the perceptual sensitivity of the operator, of the information load existing at the time, and of the intensity of the cue itself.

(5) Variability. Variability refers to the fact that some cues are static, while others are dynamic; i.e., they change over time.

Indirect Attributes: Cue/Malfunction Relationship.

(1) Relevance. Some cues are relevant in diagnosing a malfunction, while others which occur may bear no relationship to the malfunction.

(2) Diagnosticity. While relevance is an attribute that characterizes the relationship of individual cues with respect to a malfunction, diagnosticity refers to the relationship of a pattern of cues to a given malfunction. The pattern may or may not be diagnostic in identifying a malfunction.

The above cue/malfunction relationships were enumerated in an attempt to identify all possible characteristics that may be important in categorizing emergency situations. In order to simplify the description of malfunction-dependent aspects, two attributes could be used to subsume those listed above.

(1) Complexity. The pattern of cues is complex or not complex with respect to the malfunction.

(2) Congruity. There is a congruous or an incongruous relationship between the cues and the malfunction. If the relationship is congruous, the cues rather easily identify the malfunction (their pattern is reliable), whereas, if the relationship is incongruous, they represent a puzzle as to the identity of the malfunction.

2.3.2.4 Behavior-Dependent Attributes. Behavior includes overt actions, as well as decision making and other types of cognitive processes.

(1) Predictability. When behavior is said to be predictable, it is possible to specify which types of actions and decisions will be necessary to resolve the emergency situation.

(2) Likelihood. This refers to the probability that a particular sequence of actions will be utilized in response to an emergency.

(3) Programmability. Behavior is programmable or not programmable. Programmable behavior is characteristic of the actions prescribed by Boldface procedures in that entire sequences of actions can be specified and trained in advance of the emergency. Programmed behavior typically does not include complex cognitive components.

(4) Complexity. Complexity refers specifically to the complexity of the decisions that are involved in a particular situation.

(5) Criticality. Criticality, with respect to behavior, refers to the amount of time available to perform an action. Some emergency situations are highly time-critical, while others are not. This attribute refers to the behavior required rather than to the situation itself.

2.4 Initial Training Guidelines

2.4.1 Theoretical Derivations. The taxonomic structure that has been described presents a number of attributes that can be considered in the design of emergency decision training programs, and in particular, emergency training materials and mission scenarios. One attribute in particular, predictability, is a key element in describing or classifying the components of an emergency. As defined earlier, predictability refers to the specificity with which details of an emergency situation can be described and the appropriate response behaviors can be prescribed. Other attributes, of course, have implications for training program design. However, predictability, because of its central role in the representative models presented, will be explored in the remainder of this section as one guiding principle for emergency training program development.

Three major classes of emergency situations can be identified, depending on the degree of the predictability of their components. The general structure of these situations is shown in Figure 2-9. The levels of representation and the relevant components for each, are listed in the left-hand columns. For each level of representation and its appropriate components, the attribute values are specified for each of three situations. The three situations are simple labeled "predictable," "partially predictable," and, "not predictable," or situations 1, 2, and 3, respectively. In Figures 2-10, 2-11, and 2-12, the appropriate attribute values for each situation are shown separately.

Figure 2-10 defines situation 1, in which the events, the behavior, and the outcomes are predictable. In general, only single malfunctions will fall into this category, and their cues will be well-defined, recognizable, and will have high diagnosticity. These values imply that there is a simple, congruous relationship between the cues and the malfunction. When this is the case, very little decision-making is necessary at the time the malfunction is diagnosed; at the most, some problem recognition and structuring may be required. Since the event sequence is well-defined and predictable, the best decision rule can be determined at the time the event sequence is described, as can the most appropriate responses.

The cognitive structure implicit in this type of situation is that of a single template that contains all the information necessary to recognize

REPRESENTATION (MODELS)	COMPONENTS	SITUATION 1 PREDICTABLE	SITUATION 2 PARTIALLY PREDICTABLE	SITUATION 3 NOT PREDICTABLE
EVENT-RELATED FACTORS	MALFUNCTIONS CUES MALF./CUE RELATIONSHIP	(see Figure 2-10)	(see Figure 2-11)	(see Figure 2-12)
DECISION-THEORETIC FACTORS	DECISIONS DECISION RULE RESPONSE TYPE			
COGNITIVE FACTORS	COGNITIVE STRUCTURE COGNITIVE PROCESSES			
IMPLICATIONS	TRAINING REQUIREMENTS			

FIGURE 2-9.
COMPONENTS OF REPRESENTATIONAL MODELS USED TO STRUCTURE EMERGENCY SITUATIONS IN TERMS OF THEIR PREDICTABILITY

COMPONENTS	SITUATION 1
MALFUNCTIONS	SINGLE
CUES	WELL-DEFINED, RECOGNIZABLE
	HIGH DIAGNOSTICITY
	HIGH RELIABILITY
MALF./CUE RELATIONSHIP	CONGRUOUS, SIMPLE
DECISIONS	PRE-PROGRAMMED, TYPE 1
DECISION RULE	PRE-PROGRAMMED ("BEST")
RESPONSES	PROGRAMMED
COGNITIVE STRUCTURE	SINGLE TEMPLATE (MALFUNCTION AND PROCEDURE)
COGNITIVE PROCESSES	RECOGNITION, TEMPLATE MATCHING
TRAINING	BOLDFACE
REQUIREMENTS	QUICK RESPONSES
	SOME DECISION TRAINING FOR PROBLEM RECOGNITION AND STRUCTURING

FIGURE 2-10.
COMPONENTS OF SITUATION 1 EMERGENCIES

COMPONENTS	SITUATION 2
MALFUNCTIONS	A) SINGLE B) COMPOUND C) SEQUENTIAL
CUES	A) AMBIGUOUS; B) AND C) WELL-DEFINED OR AMBIGUOUS
MALF./CUE RELATIONSHIP	CONGRUOUS AND COMPLEX
DECISIONS	A) TYPE 1 B) TYPE 2 C) TYPE 3
DECISION RULE	CAN BE SELECTED
RESPONSES	PREDICTABLE, BUT NOT PROGRAMMED
COGNITIVE STRUCTURE	SEVERAL TEMPLATES
COGNITIVE PROCESSES	RECALL, TEMPLATE INTEGRATION, DISCRIMINATION JUDGMENT
TRAINING	SET, GRADUATED DECISION TRAINING
REQUIREMENTS	FLEXIBILITY (FAST "SWITCHING" OF TEMPLATES)
	DIVERSITY (GENERATION OF LOW-PROBABILITY TEMPLATES)

FIGURE 2-11.
COMPONENTS OF SITUATION 2 EMERGENCIES

COMPONENTS	SITUATION 3
MALFUNCTIONS	UNPREDICTABLE, PROBABLY COMPLEX
CUES	COMPLEX, UNCERTAIN, AMBIGUOUS
MALF./CUE RELATIONSHIP	INCONGRUOUS - SIMPLE OR COMPLEX
DECISIONS	TYPE 3 ONLY
DECISION RULE	"ASSUME THE WORST CASE" AND MINIMIZE RISK
RESPONSES	UNPREDICTABLE, NOT PROGRAMMED
COGNITIVE STRUCTURE	NO TEMPLATES
COGNITIVE PROCESSES	JUDGMENT
	CREATIVE PROBLEM SOLVING; HIGH DEGREE OF INTEGRATION REQUIRED BETWEEN ELEMENTS IN LTM
TRAINING	SET
REQUIREMENTS	TRAIN FOR RECOGNITION OF LOW-PROBABILITY EVENTS AND RELATIONSHIPS
	HIGH EMPHASIS ON PERSONAL DECISION RULE

FIGURE 2-12.
COMPONENTS OF SITUATION 3 EMERGENCIES

the malfunction and to deal with it. Thus, the process is one of recognition and of matching the correct template to the situation. The implication for training is essentially the same as that underlying Boldface procedures. The entire pattern of cues must be trained so thoroughly that the correct responses to it are immediate. In some cases, there may be some fuzzy boundaries between two or more cue patterns, so that some training in problem recognition and structuring will be required to deal with malfunctions belonging to situation 1.

Figure 2-11 lists the attribute values for the partially predictable situations. These are situations that can be foreseen, but for which decisions cannot be rigidly programmed because there are too many potential complexities that affect the decisions and the actions involved. Three types of malfunctions can belong to situation 2: single, compound, or sequential. If a malfunction is single, it has to have ambiguous cues to be categorized in situation 2. Ambiguous cues are those that suggest either no specific malfunction, or more than one malfunction, so that the cue/malfunction relationship is complex. If the malfunctions are compound or sequential, the cues can be either well-defined or ambiguous.

To some extent, compound and sequential malfunctions can be predicted, but the number of possible combinations is so great that it is not possible to present all combinations in a training course. For this reason a more generalized approach to decision training may be necessary. All three types of decision tasks (problem structuring, alternative selection, complete decision), as well as the rules for selecting the best decision, must be trained for, so that pilots will be able to evaluate applicable procedures and rules at the time of the emergency, rather than to rely on inappropriate or overly rigid prescribed responses. Desired responses are predictable in the sense that generic situations can be devised for training, but not programmed to the level of detail of a specific behavioral sequence which applies to each unique situation.

The cognitive structure underlying situation 2 emergencies consists of several templates, or of representational systems with overlapping elements, so that feature recognition and integration of the information from several sources is necessary. This requires more active recall than the simple recognition and machine of situation 1 emergencies. Because the set of potential situations belonging to this category is very large, training materials must be carefully structured and carefully controlled to ensure that all essential factors are included, and that they are graduated with respect to their difficulty.

Figure 2-12 presents the conceptualization of unpredictable situations. The malfunctions are unpredictable and probably complex. The cues may be complex, but they are certainly ambiguous because if the malfunction cannot be foreseen, the cues will probably not display a recognizable pattern. If the cue pattern is recognizable, it may be misleading, so that the relationship of the cues to the malfunction will be incongruous. In these cases, only complete decisions (Type 3) will be appropriate, and an appropriate decision rule is to minimize the risk and to assume the worst possible outcome. No templates will be available to deal with this situation, but a high degree of integration between disparate representational systems will be required. Effective responses may be compared to creative problem solving, namely, to apply old solutions to problems that have never been encountered before, or to induce the occurrence of uncommon responses. The training requirements are similar to those for situation 2 emergencies, but they must emphsize this added creative aspect--practice in generating low-probability events and procedures. This emphasis might be achieved by presenting simulated situations that require unusual solutions and by reinforcing the use of such solutions.

2.4.2 Review of Training Implications. The three classes of emergency situations described above were categorized according to their predictability because of the central role this attribute plays in the representations and taxonomic structure developed for aircraft emergencies. Figure 2-13 is a brief overview of the three classes of emergencies in terms of the taxonomic structure together with training implications for each class.

For Situation 1, Boldface-like training appears to be relevant, assuming the presence of time and safety criticality. These emergencies involve straightforward relationships between cues and malfunctions information processing requirements are low, and response procedures are known and programmable.

For Situation 2, explicit training in decision techniques is recommended since a less predictable set of circumstances and responses is involved than in Situation 1. Cues are complex and/or ambiguous with respect to identifying malfunctions. More information processing is required, and responses cannot be fully programmed ahead of time. Training techniques which emphasize integration of several cognitive representational systems appear to be recommended.

In Situation 3, cues can be complex, ambiguous, and perhaps misleading. Responses are not programmable ahead of time and extensive deliberation may be necessary to diagnose the situation and develop an appropriate response. Training for these emergencies must address the ability to integrate disparate representational systems and to account for low probability events and relationships. Development of personal decision rules, in which the pilot establishes techniques for dealing with manageable approximations of complex situations, also appears to be recommended.

	SITUATION 1	SITUATION 2	SITUATION 3
MALF./CUE RELATIONSHIP	CONGRUOUS SIMPLE	COMPLEX CONGRUOUS	INCONGRUOUS
RESPONSES	PROGRAMMED	NOT PROGRAMMED. PREDICTABLE	NOT PROGRAMMED, NOT PREDICTABLE
COGNITIVE PROCESSES	TEMPLATE MATCHING	TEMPLATE INTEGRATION	CREATIVE PROBLEM SOLVING
TRAINING IMPLICATIONS	BOLDFACE QUICK RESPONSES SOME DECISION TRAINING FOR PROBLEM RECOG-NITION AND STRUCTURING	SET, GRADUATED DECISION TRAINING FLEXIBILITY (FAST "SWITCHING" OF TEMPLATES) DIVERSITY (GENERATION OF LOW-PROBABILITY TEMPLATES)	SET, TRAINING FOR RECOGNITION OF LOW PROBABILITY EVENTS AND RELATIONSHIPS HIGH EMPHASIS ON PERSONAL DECISION RULE

FIGURE 2-13.
OVERVIEW OF INITIAL TRAINING GUIDELINES

The taxonomy represents a framework for combining theoretical implications into a set of hypotheses relevant to decision making and cognitive behavior in emergency situations. The three situations proposed here are differentiated on the basis of the degree of predictability of the types of responses necessary to deal with various emergency conditions. Other attributes may also have differential implications which need to be explored and applied to the development of training guidelines.

3. EXPERIMENTAL STUDIES

3.1 Introduction

The overall objective of the experimental work of Year 2 was to identify and specify relevant decision-related *cognitive factors* associated with aircraft emergency situations. The ultimate goal is to facilitate the design of synthetic learning environments for the training and assessment of high performance decision skills.

This chapter describes the four experimental studies carried out during Year 2 and presents selected findings from them. Study 1 was conducted to isolate variables for study and to test and refine the basic experimental procedures. Study 2 was designed to determine how individual or groups of cues (events which might signal aircraft malfunctions) are perceived in terms of basic cognitive dimensions (cue attributes) which are hypothesized to underlie emergency decision-making behavior. Study 3 was an attempt to validate the underlying assumption that cues or cue patterns, hypothesized to be significantly related to decision making, did, in fact, produce differences in decision making behavior. Study 4 was designed to expand on Study 2 by further refining the cues and cue attributes examined and by including additional situational factors which would be of significance in the performance environment.

3.2 Study 1

This exploratory study was conducted with the assistance of nine experienced F-4 pilots from Miramar Naval Air Station who served as subjects. A number of mini-scenarios, consisting of various potential malfunctions and situational factors, were used as test materials and presented to each subject for solution. Protocol analyses were carried out to determine whether subjects were able to identify and describe the choices, considerations, and decision phases involved in each scenario reviewed. In general, it was found that the pilots were not able to identify or describe their decision processes with any degree of specificity, although they easily developed a recognition of the problem presented and an acceptable course of action. It was also found that the pilots varied in their interpretation of the cues (events or stimuli) which were presented, namely in terms of the importance of such cues and the severity of the possible malfunctions which might exist.

A number of useful outcomes were derived from this study. First, a very detailed scenario for the problems surrounding stall/stagnation was developed, based on the pilots' responses and interactions with the interviewer. Second, conversations with the pilots indicated that certain aspects of the cues or events making up the mini-scenarios were important, but pilots varied in terms of the cue aspects that they emphasized. This finding suggested that a more basic examination of certain aspects or attributes of the cues involved in emergency situations would be profitable. Finally, the inability of pilots to identify or describe their decision approach to any degree of detail suggested that a more experimental approach was required. The results of this first study were used to design Study 2 and Study 3 described below.

3.3 Study 2

Two basic questions were examined in this study. First it was of interest to see whether and how pilots perceived differences in cues and cue patterns in terms of six underlying attributes which are hypothesized to be related to decision making performance. Second, it was of interest to see whether and how pilots' perceptions of the cues in terms of the underlying attributes varied as a function of cue combination.

Six cue attributes were considered in this study on the basis of pilots' responses in Study 1. These cue attributes, which were selected from among the larger set described in the theoretical analysis of Year 1 of this project, are:

(1) Salience--recognizability of the cue.

(2) Diagnosticity--degree to which the cue helps isolate a particular malfunction.

(3) Variability--fluctuation or variation in the nature of the cue over time (static vs. changing).

(4) Predictability--degree to which cue clearly signals future events and outcomes.

(5) Time Criticality--degree to which cue indicates that a rapid response is required.

(6) Safety Criticality--degree of danger indicated by the cue.

It was hypothesized that the first four attributes are directly related to separate phases of decision-making performance as shown below:

Attribute	Decision Phase
Salience	Problem Recognition
Diagnosticity	Problem Structuring
Variability/Predictability	Alternative Selection

The last two attributes, time criticality and safety criticality, were assumed to be related to task stress or difficulty, which would affect performance across all phases of decision behavior.

In Study 2, 10 different experienced F-4 pilots from Miramar Naval Air Station served as subjects. All pilots were presented with 24 cues or patterns of cues to rate in terms of the six attributes described above. Figure 3-1 presents the cues used. The seven cues were chosen as being

1 - Mild Engine Vibration

2 - Strong Engine Vibration

3 - Mild Explosion (Thump)

4 - Loud Explosion

5 - Minor Engine Surge

6 - Significant Engine Surge

7 - Engine Seizure

8 - Attitude Indicator Failure

9 - Tailhook Light On

FIGURE 3-1.
CUES USED IN STUDY 2

relevant to a malfunctioning engine according to Section III of the F-4 Manual. Two additional cues, attitude indicator failure and tailhook light on, were chosen as irrelevant cues for this malfunction, and were included for purposes of comparison. Fifteen additional patterns of cues were made up using the nine single cues in representative configurations with the aid of an experienced pilot. All subjects were asked to rate each cue or cue pattern separately on a scale from 0 to 10 for each attribute, the scale being defined individually for each attribute. A counterbalanced order of presentation was employed. In rating each cue or cue pattern, pilots were asked to assume that flight conditions were good visibility, high altitude, and fast speed.

Figure 3-2 presents the distributions of the mean ratings for the cues/cue patterns on each of the six attributes considered by the pilots. Review of Figure 3-2 indicates that some discrimination among cues/cue patterns occurred for all attributes, with the greatest spread occurring for time and safety criticality and the least for predictability and variability. Inspection of the standard deviations of the ratings indicated that inter-rater agreement was highest (low standard deviations) for salience and safety criticality, and lowest for predictability, variability, and time criticality.

These findings support the results of Study 1 which suggested that pilots disagree to some extent on the significance of various aspects of events (cues) in emergency scenarios. More importantly, however, the ratings and standard deviations provide a means to estimate the impact of individual cues on a given aspect of the decision task. For example, a loud explosion (Cue 4) was perceived as high on problem recognition and safety and time criticality, but only moderate with respect to problem structuring and alternative selection. To the extent that reliable and discriminable ranking such as these can be determined for each cue attribute of interest, the design of emergency scenario components which permit differential treatment of the different phases of decision making should be facilitated.

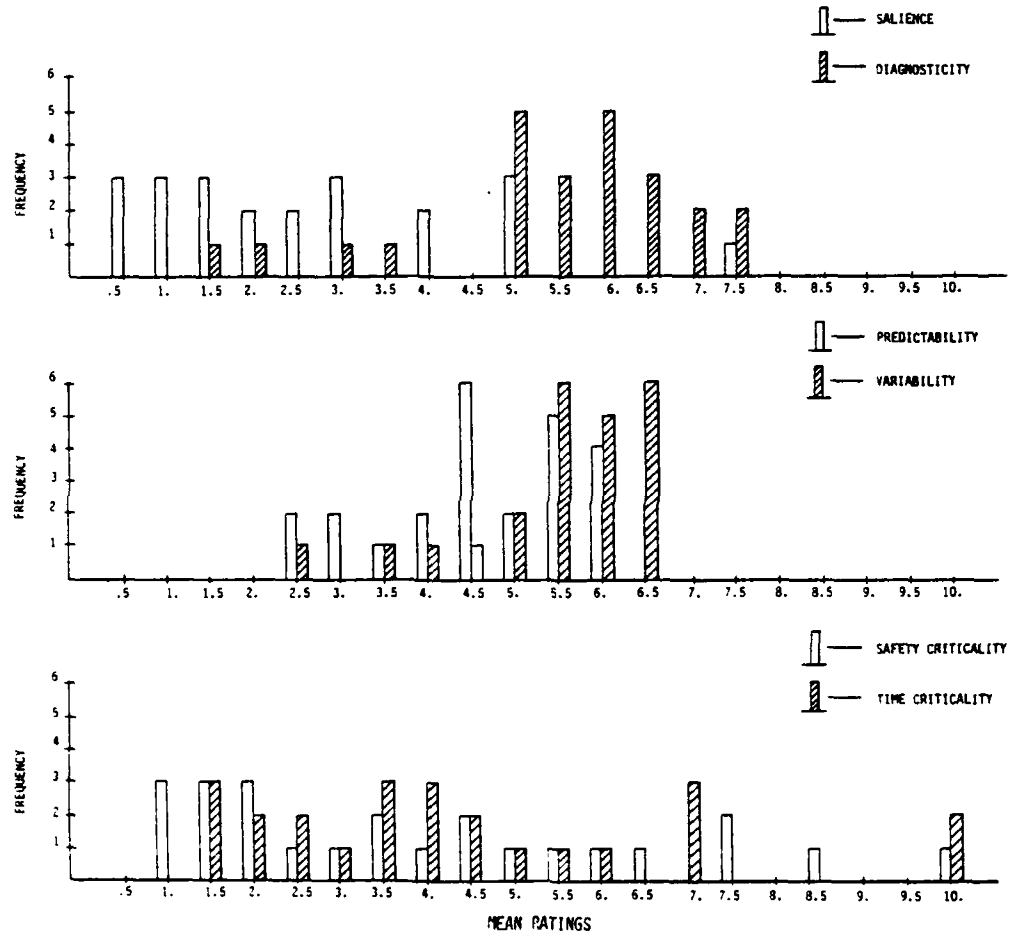


FIGURE 3-2.
DISTRIBUTIONS OF MEAN RATINGS OF CUES
AND CUE PATTERNS FOR SIX ATTRIBUTES

With respect to the question of cue interaction, it was of interest to examine how attribute ratings varied depending on the component cues included in a pattern. Figure 3-3 demonstrates how ratings of cue patterns shift from the rating for an individual cue when presented alone. In Figure 3-3, the ratings for all cue patterns containing engine seizure (Cue 7) are shown. Addition of from one to three other cues has little effect on pilots' ratings of salience, time criticality, of safety criticality. Ratings of diagnosticity and variability become much higher, suggesting that the cue loses impact (is less diagnostic and more variable) when clustered with these confounding or distracting events.

Figure 3-4 summarizes the information obtained with respect to three cues, showing shifts in ratings from the single cue rating in absolute terms. The most distinct cue (loud explosion) is affected least in terms of shifts in ratings when other cues are added, while the least distinctive of the three shown in Figure 3-4, minor engine surge, is most affected.

Figure 3-5 summarizes some additional data from Study 2, namely pilots' perceptions of the effects of various situational factors on cue ratings. The responses shown suggest that complicating environmental or situational factors might alter the responses (and decision behavior) elicited by the cues considered in this study, as might be expected on an intuitive basis. Study 4, to be described below, was intended to consider the impact of such situational factors in greater detail.

3.4 Study 3

This study was intended to test whether cues/cue patterns assumed to impact on the various phases of decision behavior produce differential effects on such behavior as evidenced by pilot's responses to mini-scenarios containing such cues/cue patterns. Specific hypothesis were:

DECISION COMPONENT	PROBLEM RECOGNITION	PROBLEM STRUCTURING	ALTERNATIVE SELECTION		STRESS	
ATTRIBUTES	SALIENCE	DIAGNOSTICITY	PREDICTABILITY	VARIABILITY	SAFETY CRITICALITY	TIME CRITICALITY
SINGLE CUE RATING	1.5	2.5	3.0	2.0	2.0	2.5
SHIFTS						
+4.5		(1,4,9)				
+4.0						
+3.5				(1,4,9)		
+3.0		(4)(1,4) (1,5,9)		(1,4)		
+2.5			(1,4,9)	(1,3,5)		
+2.0		(5,1) (1,3,5)		(5,1) (4)		
+1.5				(5,1,9)		
+1.0			(4)(1,4)			
+ .5			(5,1,9)			(5,1)(5,1,9)
0						
- .5	(5,1)(5,1,9)		(1,5)(1,3,5)		(5,1)(5,1,9)	(1,3,5)
-1.0	(1,3,5)				(1,4)(1,3,5)	
-1.5	(4)(1,4) (1,4,9)				(4)(1,4,9)	(4)(1,4)(1,4,9)

FIGURE 3-3.
SHIFTS IN RATINGS PRODUCED BY ADDING CUES TO CUE 7 (ENGINE SEIZURE)

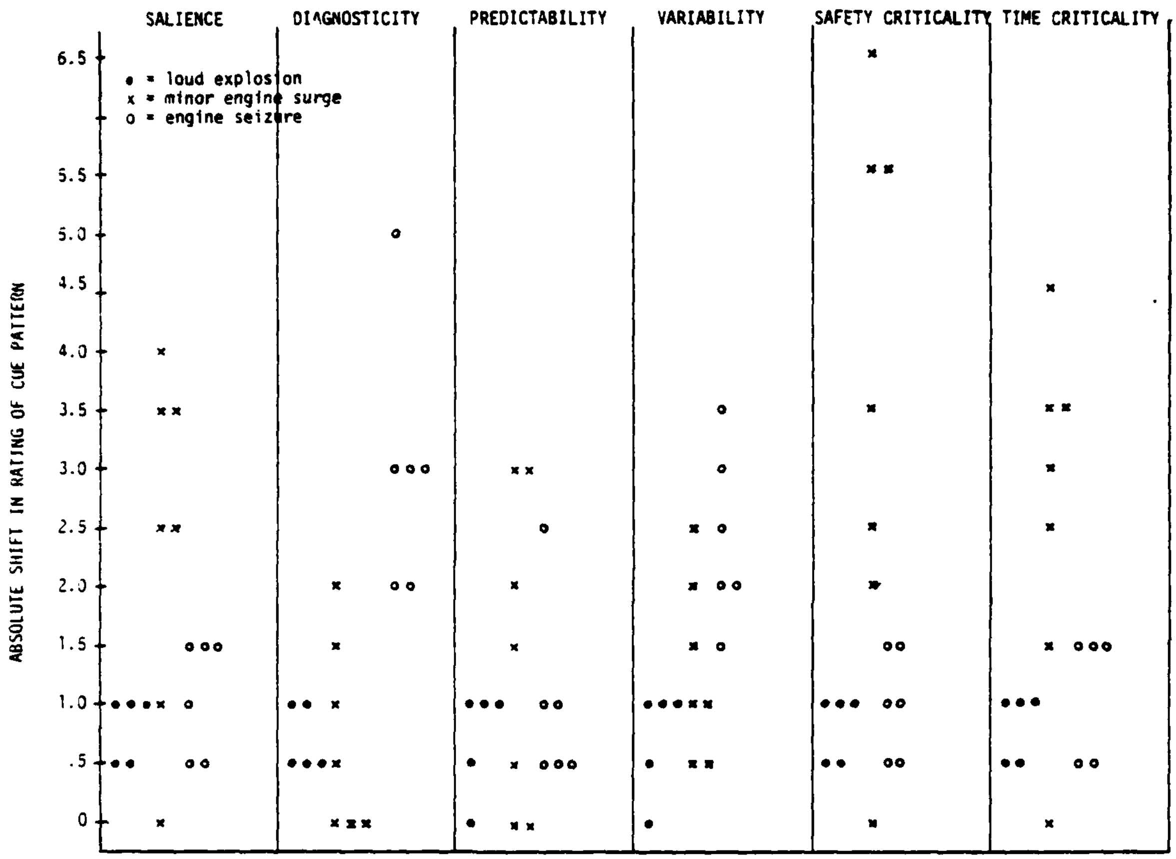


FIGURE 3-4.
SHIFTS IN ATTRIBUTE RATINGS FOR THREE CUES
WITH ADDITION OF OTHER CUES TO CUE PATTERN

SITUATIONAL FACTORS

CUE ATTRIBUTE	BAD WEATHER	LOW ALTITUDE FAST FLYING	LOW ALTITUDE SLOW FLYING	COMBAT
Salience	80%	40%	40%	80%
Diagnosticity	29%	29%	29%	71%
Predictability	29%	14%	29%	43%
Variability	33%	33%	33%	50%
Safety Criticality	100%	67%	67%	83%
Time Criticality	86%	71%	71%	86%

FIGURE 3-5.
PILOT PERCEPTIONS OF IMPACT OF SITUATIONAL FACTORS
ON CUE ATTRIBUTES

(1) That cue patterns rated as having high impact with respect to diagnosticity (problem structuring) or variability and predictability (alternative selection) would result in fewer responses than those rated as low in impact.

(2) That cue patterns with high ratings for time and safety criticality would elicit fewer responses than patterns rated low on these attributes.

In the first case, it is assumed that highly diagnostic and low variability/high predictability cues signify clear-cut or easy decision problems. In the second case, it is assumed that highly critical situations will evoke trained (Boldface) responses and that the introduction of stress will tend to reduce the range of possible malfunctions and alternative actions considered.

To test these hypotheses, six experienced F-4 pilots from Miramar Naval Air Station were each presented a set of situations or mini-scenarios formed by selecting cue patterns from Study 2 which fell at the extremes for the attributes being considered. Four cue patterns were used to elicit responses from the pilots regarding the possible problems which might exist (problem structuring). Four patterns were also used to elicit responses regarding possible courses of action (alternative generation).

The results for problem structuring confirmed the hypothesis, namely that more possible malfunctions were identified for the cue pattern rated low on diagnosticity that the pattern rated high. Also, for patterns high and low on diagnosticity, the addition of task stress in the form of cues rated high on time and safety criticality resulted in a 50% reduction in the number of malfunctions identified.

With respect to alternative selection, the hypotheses were not confirmed. There was little variation, in general, in the number of alternative actions generated regardless of level of task stress or predictability/validity. One explanation may be that the cue attributes examined do not in fact impact on alternative selection. The absence here of an effect for task stress, however, which had an impact on problem structuring, argues against this interpretation. An alternate and more like explanation is that pilots are trained to think in terms of certain preferred actions for a given situation and are unlikely to generate multiple alternatives without special effort. The relationship hypothesized between cue attributes and alternative selection should be tested further using a revised experimental procedure which is more sensitive to the possible effects of cue attributes.

3.5 Study 4

3.5.1 Approach. Study 4 is an outgrowth of the activities and findings of Studies 1, 2, and 3. Studies 2 and 3 had been designed to test the hypothesis that a link between the objective attributes of the event model (specifically, the attributes of malfunction-related cues) and the cognitive processes specified by the decision model, could be specified. Such a link would make it possible to empirically manipulate the difficulty level of the decision components for inclusion in synthetic learning environments. Specifically, it was hypothesized that ratings could be obtained on six different attributes of cues or cue patterns and that these attributes had differential impact on the three decision making phases of problem recognition, problem structuring, and alternative selection. The ratings were obtained in Study 2, and the actual hypothesis tested in Study 3. The results obtained were promising and suggested that this was a feasible approach to the goal of empirically quantifying the difficulty level of the three decision making phases.

The assumption underlying Study 2 was that meaningful differences in the difficulty level of situations could be obtained by asking experts to rate various cues and cue patterns on the attributes of salience, diagnosticity, predictability, variability, safety criticality and time criticality. This basic assumption also underlies Study 4. In fact, the mechanics and the procedures of Study 2 and Study 4 were identical, but the contents and the focus of investigation differed. Specifically, if differences in the perceived value of the attributes are to be used in generating synthetic learning environments that are controllable with respect to the difficulty of the decision task, several factors need to be investigated concerning the characteristics and the interrelationships of the cues making up the situations of mini-scenarios.

In Study 2, judgments were obtained using mini-scenarios consisting of a small set of cues, all pointing to a single malfunction (engine problem), plus two irrelevant cues. No irrelevant cues per se were used in Study 4. Rather, the emphasis was on additional issues that had not been considered previously. Specifically, Study 4 was intended:

(1) To test the hypothesis that situational factors (weather variables and flight phases) can be treated in a similar fashion to malfunction-related cues in quantifying the difficulty of decision tasks.
(2) To investigate the effects of using cues suggestive of more than a single malfunction.
(3) To determine differences in the judgments of experienced and inexperienced pilots with respect to cue attributes.
(4) To replicate and extend the data base established in Study 2.

As noted, it was hypothesized that situational factors could be included in the mini-scenarios and treated in the same manner as malfunctioning cues (malfunction cues are defined as those cues that specifically indicate a system malfunction of some sort). Second, since all cues in Study 2 pointed to a single malfunction, it was important to design Study 4 to include cues belonging to different malfunctions.

For Study 4, seven malfunction cues and four situational cues shown in Figure 3-6 were selected and combined into patterns. The malfunction cues were taken from the Emergency Procedures Section of the F-4 Manual in such a way that they formed patterns that were either congruous or incongruous. A congruous pattern was defined as a pattern having cues that were all described as belonging to a particular malfunction; for an incongruous pattern, cues that belonged to several different malfunctions were combined. The two new situational cues which were considered, which represented deviations from "ideal" flying conditions (good visibility and flying high and fast). were poor visibility and flying low and fast.

An additional factor of interest was pilot experience level as it affected basic judgments of cues and cue patterns. Two levels of pilot experience were utilized in order to determine whether limited or extensive experience with the selected aircraft (the F-4) affected judgments of decision-related cue attributes.

The malfunction-related and the situational cues were combined into 77 mini-scenarios (cue patterns) and presented for rating to two group of pilots at Luke AFB, AZ. The two groups were made up of 13 experienced F-4 pilots (average number of hours in the F-4, approximately 800) and 11 inexperienced pilots (average number of F-4 hours, approximately 20) who were asked to rate the situations on the six attributes used in Study 2, namely salience, diagnosticity, variability, predictability, time criticality and safety criticality. Ratings varied from zero (low) to 10 (high) for each attribute. A cue rated low on salience would be considered hard to recognize, and so on. Each pilot rated half the cue patterns on three attributes and half on the other three attributes. This yielded between five and seven judgments on each situation for each attribute in the two groups. Means and standard deviations of those judgments were obtained and the results were used for most of the subsequent analyses.

CUE

1 - ERRATIC FUEL GAUGE

2 - RH GENERATOR LIGHT ON

3 - MILD GENERAL VIBRATION

4 - STIFF THROTTLES

5 - STUCK THROTTLES

6 - LOSS OF RPM

7 - FUEL FUMES IN COCKPIT

A - GOOD VISIBILITY

B - POOR VISIBILITY

C - FLYING HIGH AND FAST

D - FLYING LOW AND FAST

FIGURE 3-6.
CUES USED IN STUDY

3.5.2 Results. Figure 3-7 presents the average ratings by pilot experience level for the six attributes considered. The distributions indicate the discriminability of the ratings of the attributes. In Study 2, the ratings were much more bunched together than here, suggesting that the cues and cue combinations used did result in a wider range of perceptions of the six attributes across the various cues/cue pattern combinations. This finding indicates that the attempt to sample a more varied set of cues was successful.

Figure 3-8 presents the distributions of standard deviations for ratings of cues/cue patterns by experienced and inexperienced pilots. Although for several cue patterns, agreement among raters was low on one or more attributes, the distributions reveal than in general agreement among raters was quite good. There were no cases, except for variability ratings, where the standard deviations of the experienced pilots were clearly smaller than for the inexperienced pilots.

These preliminary findings (Figures 3-7 and 3-8) do not differentiate strongly between the two experience groups. This is contrary to the initial hypothesis, but may be explained in part by the fact that even the experienced group had had a great deal of prior experience in T-37 and T-38 fighter aircraft. The situations selected for this study were system specific in that the cues used were taken from the Dash-1 of the F-4 Manual; however, these were situations that could also occur in other fighter aircraft, thus adding to the level of experience that new F-4 pilots brought to the study. Also, there was a certain amount of overlap between the experience levels of the two groups, in that some of the inexperienced pilots already had had up to 100 hours of experience in the F-4.

An additional analysis which compared only the most experienced *versus* the least experienced pilots in the two groups failed to yield any significant differences on overall ratings of attributes across cues. There were some

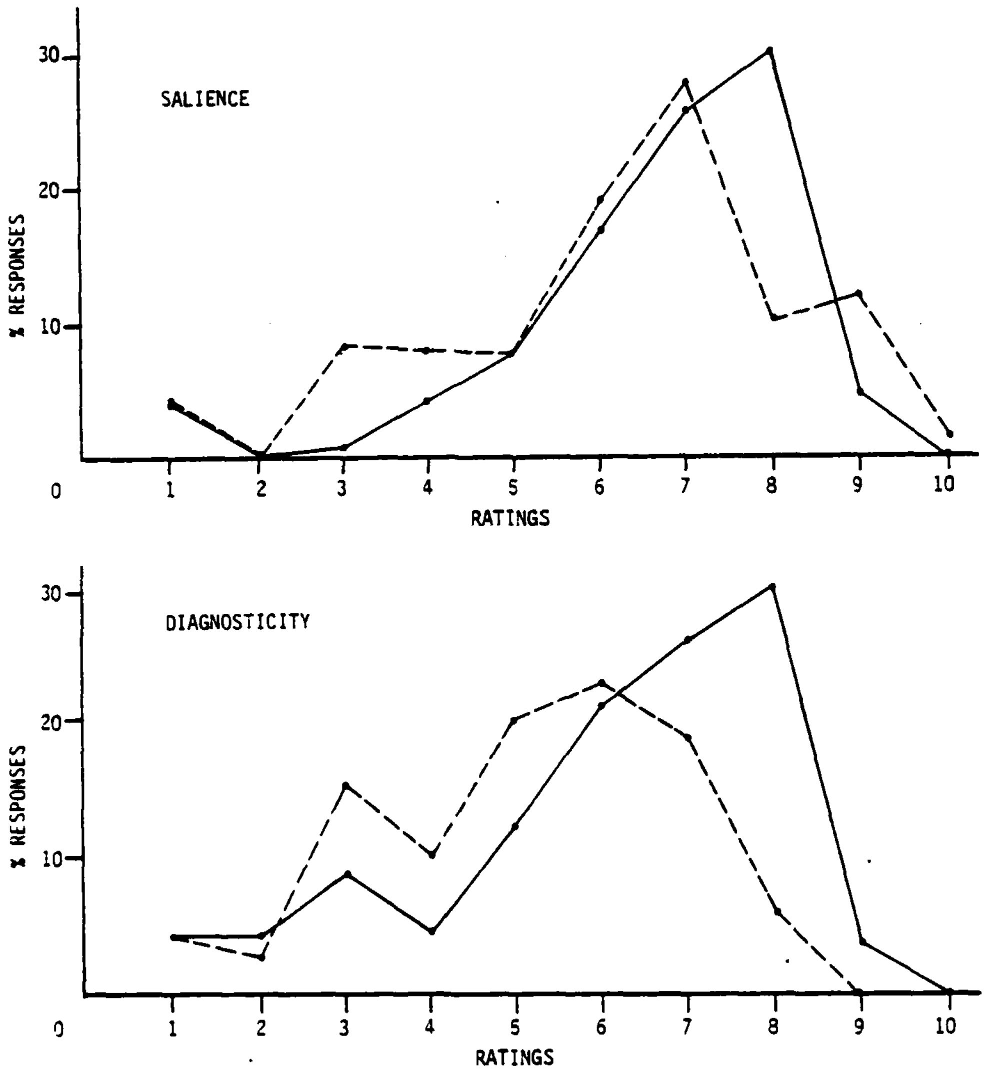


FIGURE 3-7.
DISTRIBUTIONS OF MEAN RATINGS FOR ATTRIBUTES

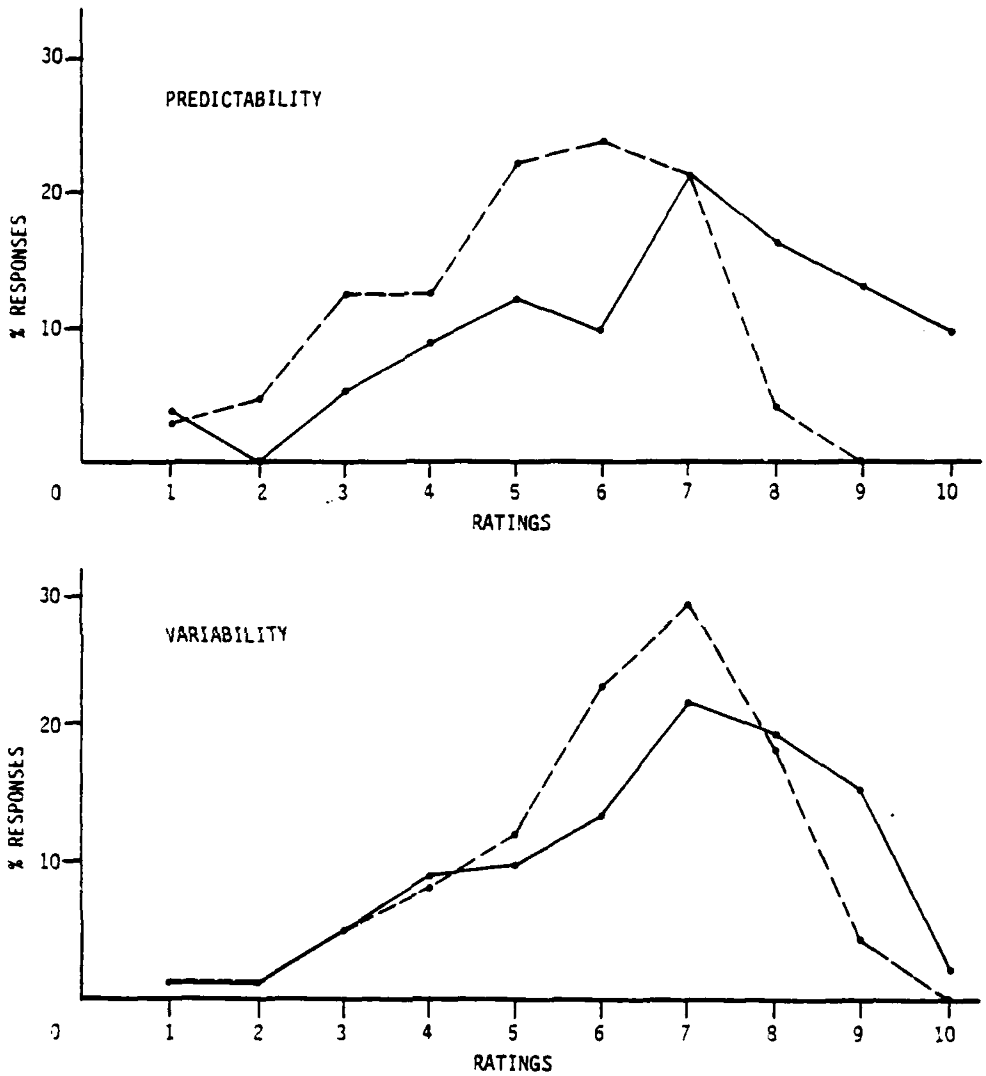


FIGURE 3-7. (CONT.)

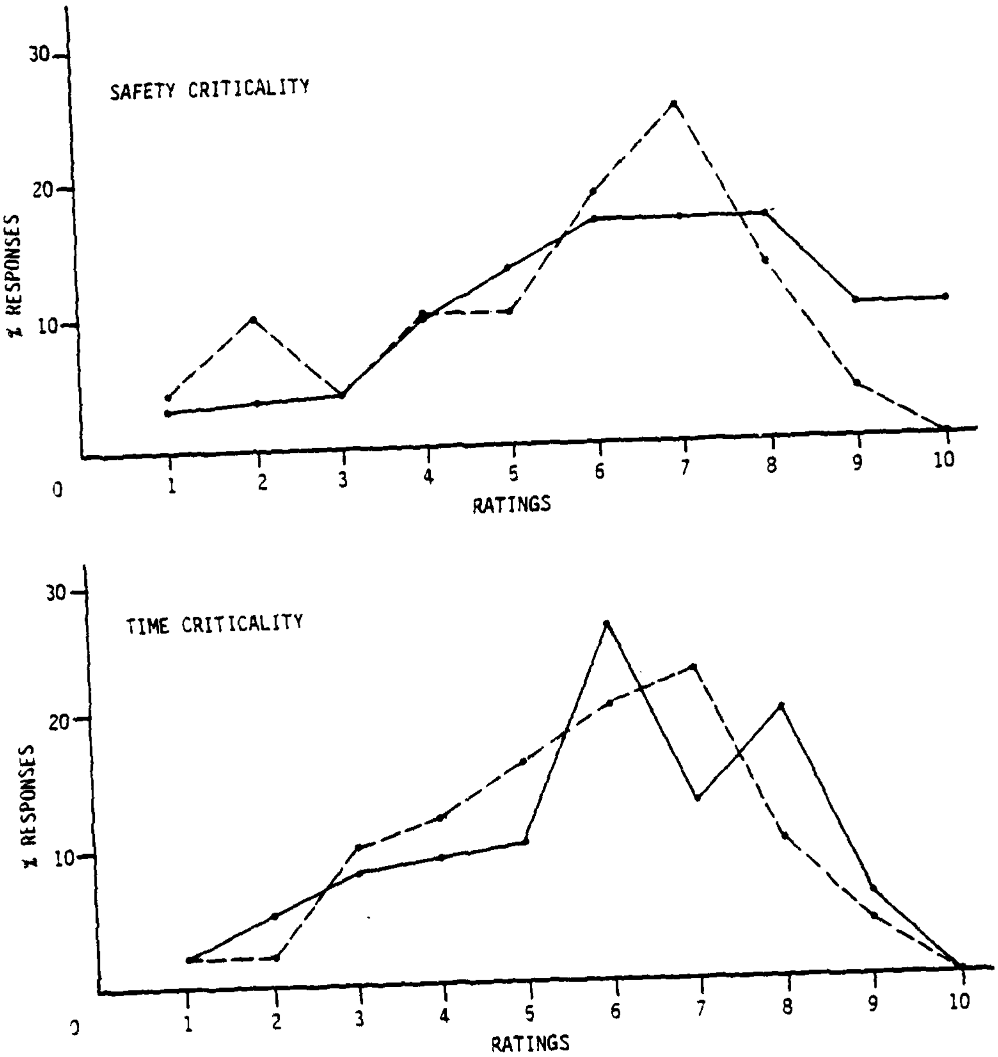


FIGURE 3-7 (CONT.)

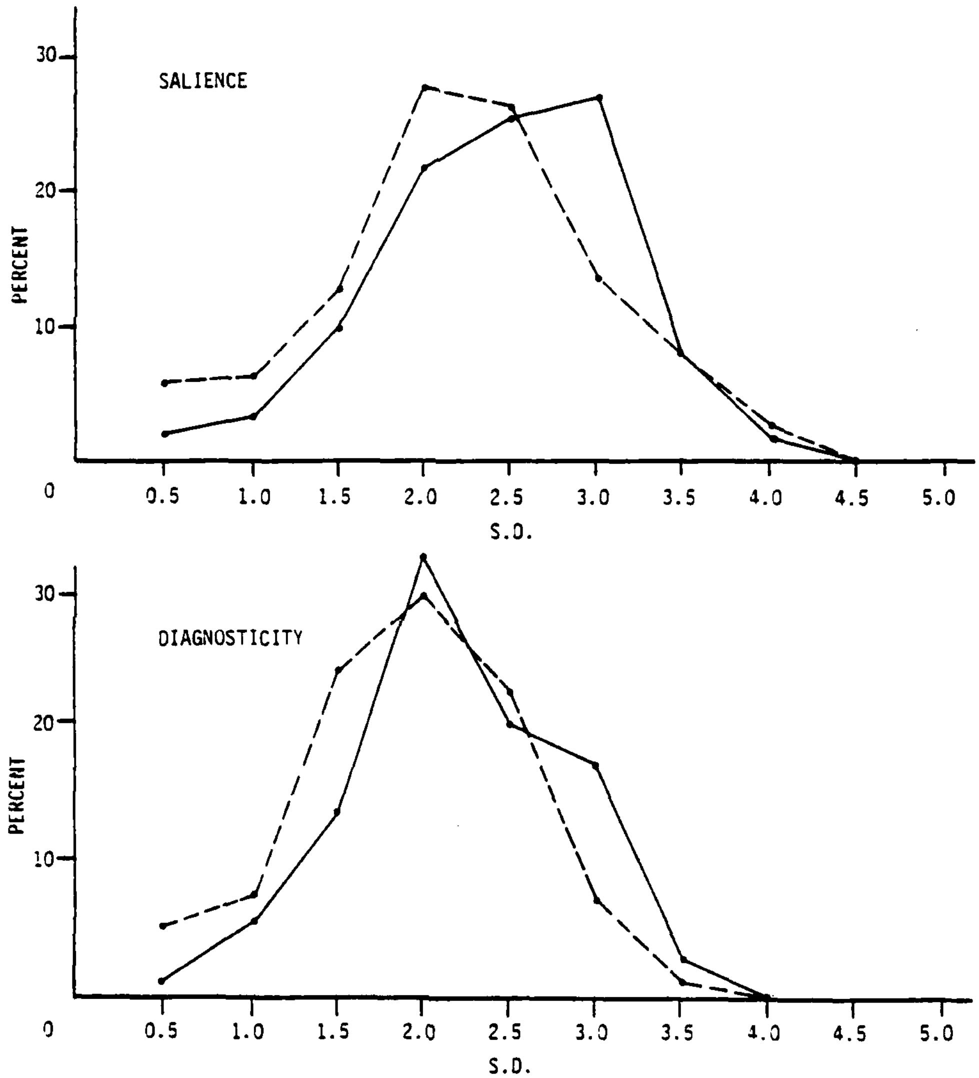


FIGURE 3-8.
DISTRIBUTIONS OF STANDARD DEVIATIONS FOR ATTRIBUTE RATINGS

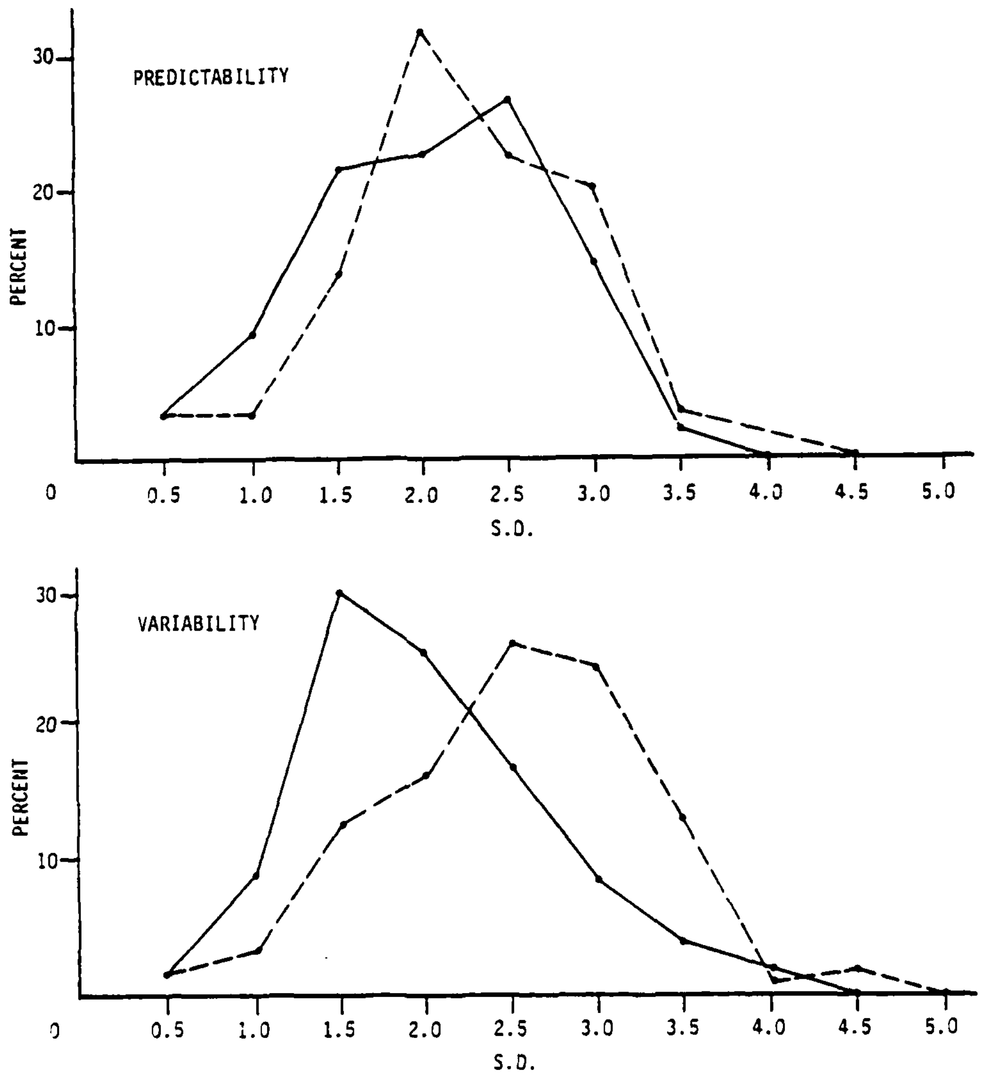
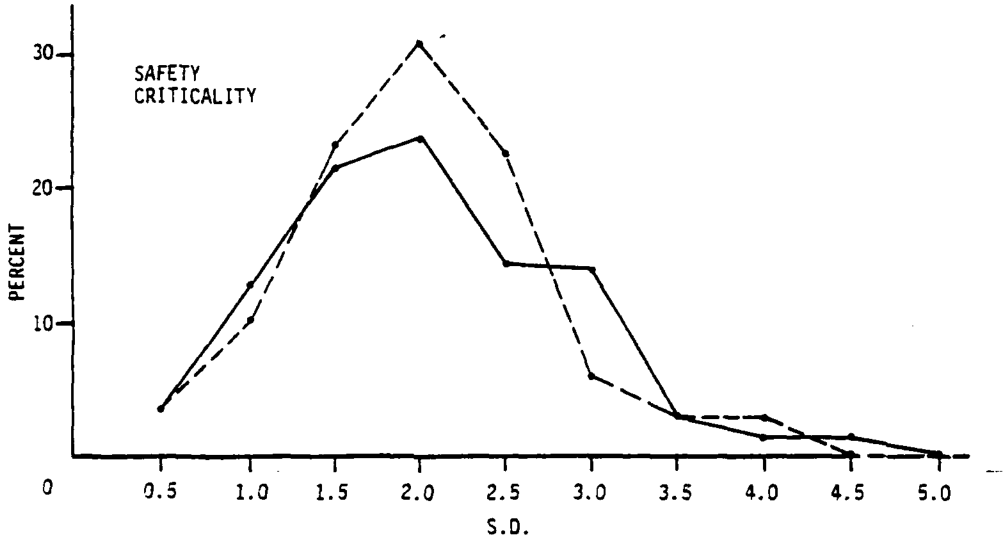


FIGURE 3-8. (CONT.)



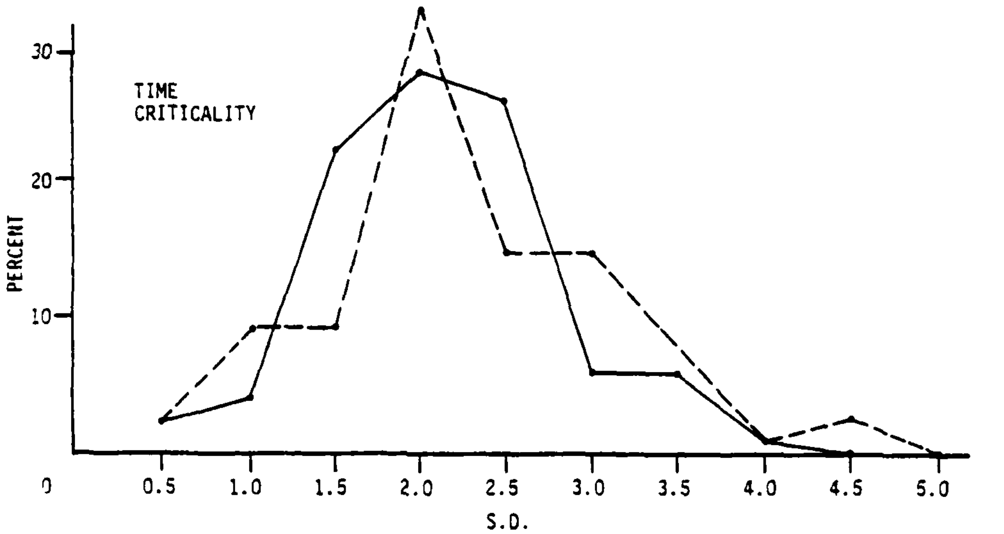


FIGURE 3-8. (CONT.)

suggested differences between experience groups when the ratings were analyzed by specific cue patterns, and these will be briefly discussed below:

Figure 3-9 summarizes the impact of adding additional cues to a single malfunction cue. The figures illustrate on a cumulative basis, across attributes, how the ratings on a particular cue shift when other cues are added to it to form a cue pattern. For each curve, the steeper the slope and the closer it remains to zero, the less the impact that addition of other cues has on the rating. Conversely, the steeper the slope and closer to zero, the greater the dominance the cue has in its pattern. The results for experienced pilots indicate less shift in ratings for cues 3, 4, 5, and 7 than was found with inexperienced pilots. For the other three cues, only the ratings for Cue 6 indicated that the inexperienced pilots had more stable ratings. Comparisons of cues indicate that Cue 1 showed the most shift, while Cues 5 and 7 were the most dominant or showed the least shift.

Figure 3-10 similarly presents shift data for each attribute by pilot experience level, averaged over cues. The ratings for salience shows the least shift, while those for time and safety criticality show the most. As was the case for the shift data when presented by cue, the ratings of experienced pilots tended to be slightly more stable than those of the inexperienced pilots. Ratings by experienced pilots for salience showed much less shift than ratings by inexperienced pilots. Conversely, the ratings of variability were more stable for inexperienced pilots.

Figures 3-11 through 3-16 provide a more detailed look at the underlying judgment data obtained, since shifts in judgment are shown for individual cues and attributes, by pilot experience level. These figures present the basic judgment data which might be used to construct elements of emergency decision problems. Selection of particularly dominant cues may be made from these data and used to construct cue patterns which result in consistent or misleading cue patterns, as desired.

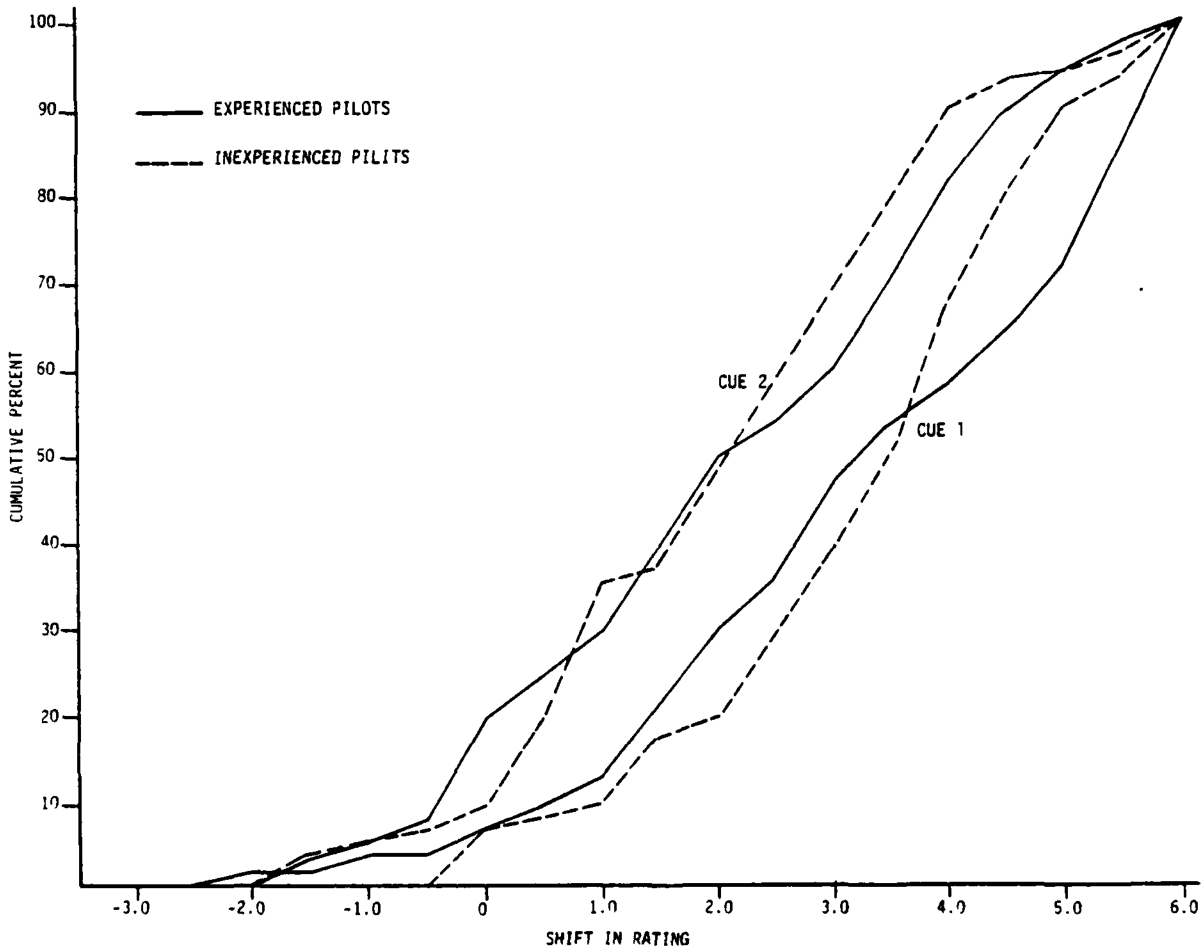


FIGURE 3-9.
CUE DOMINANCE FOR TWO PILOT GROUPS

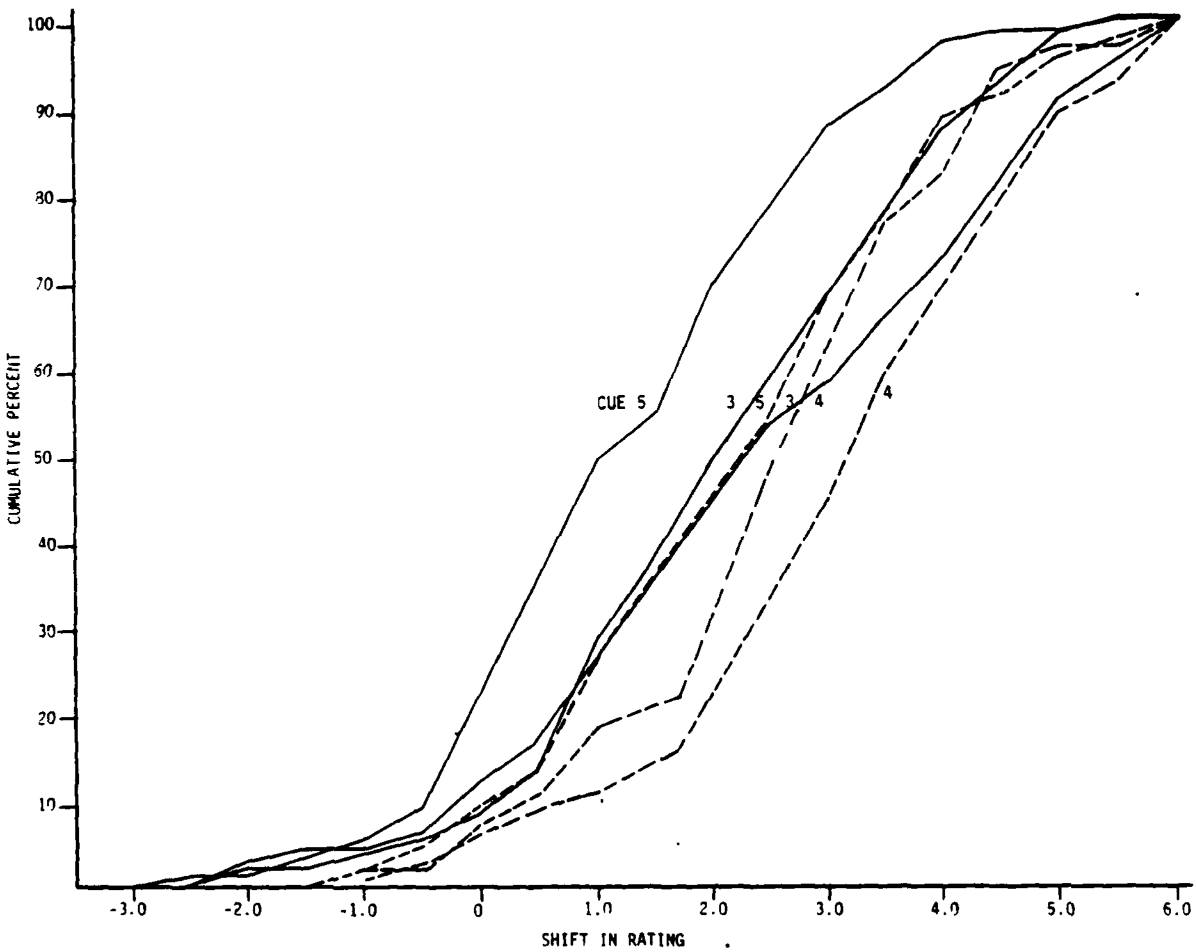


FIGURE 3-9. (CONT.)

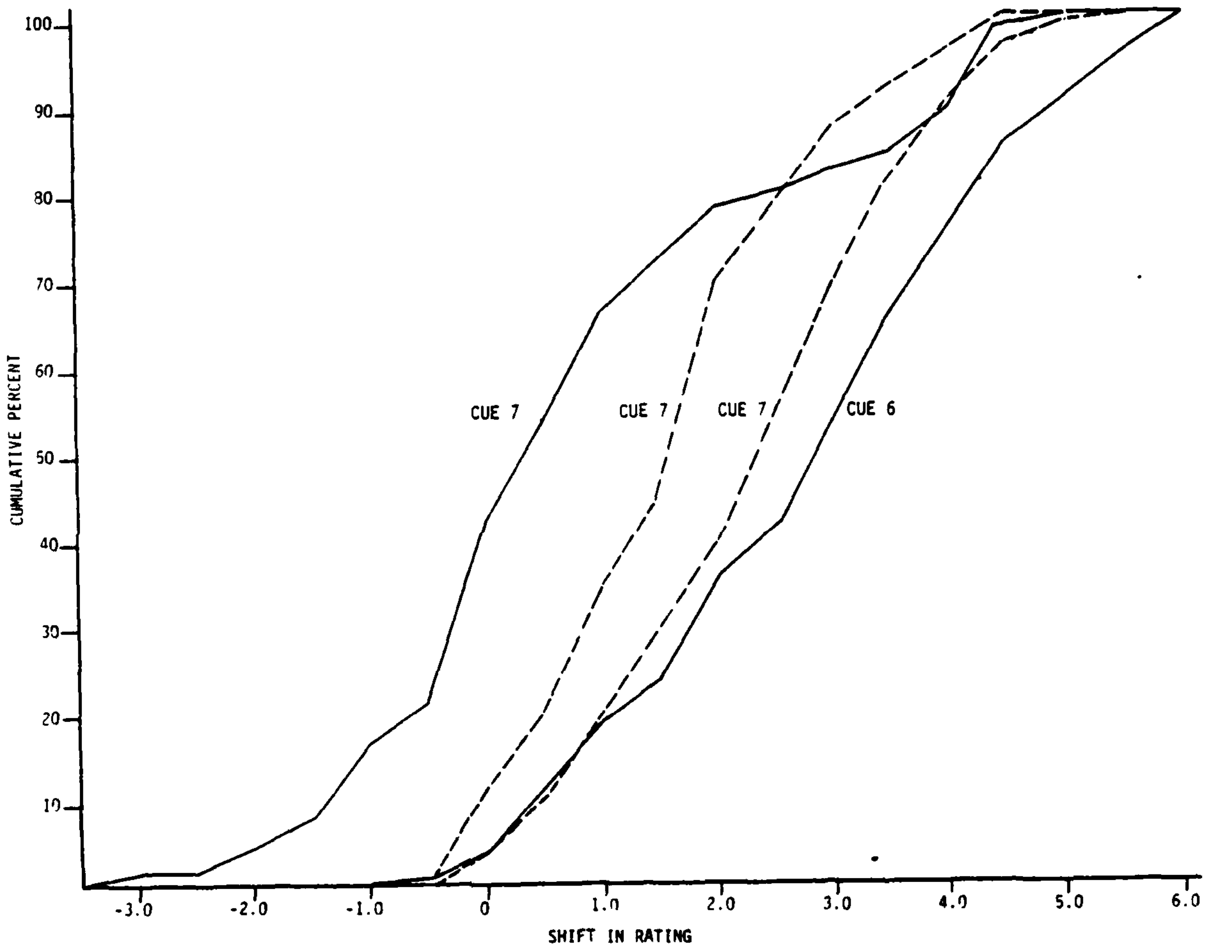


FIGURE 3-9. (CONT.)

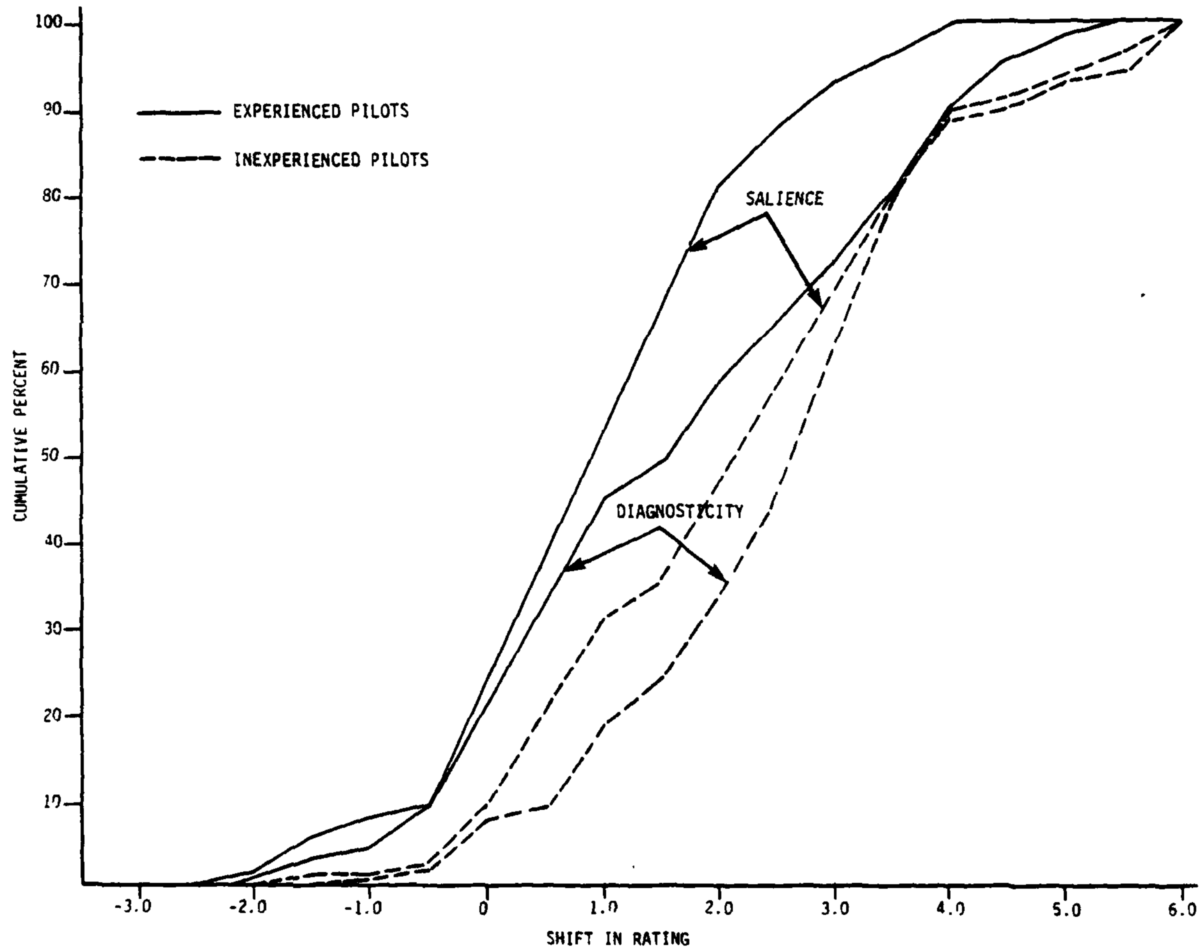


FIGURE 3-10.
SHIFTS IN RATINGS OF ATTRIBUTES

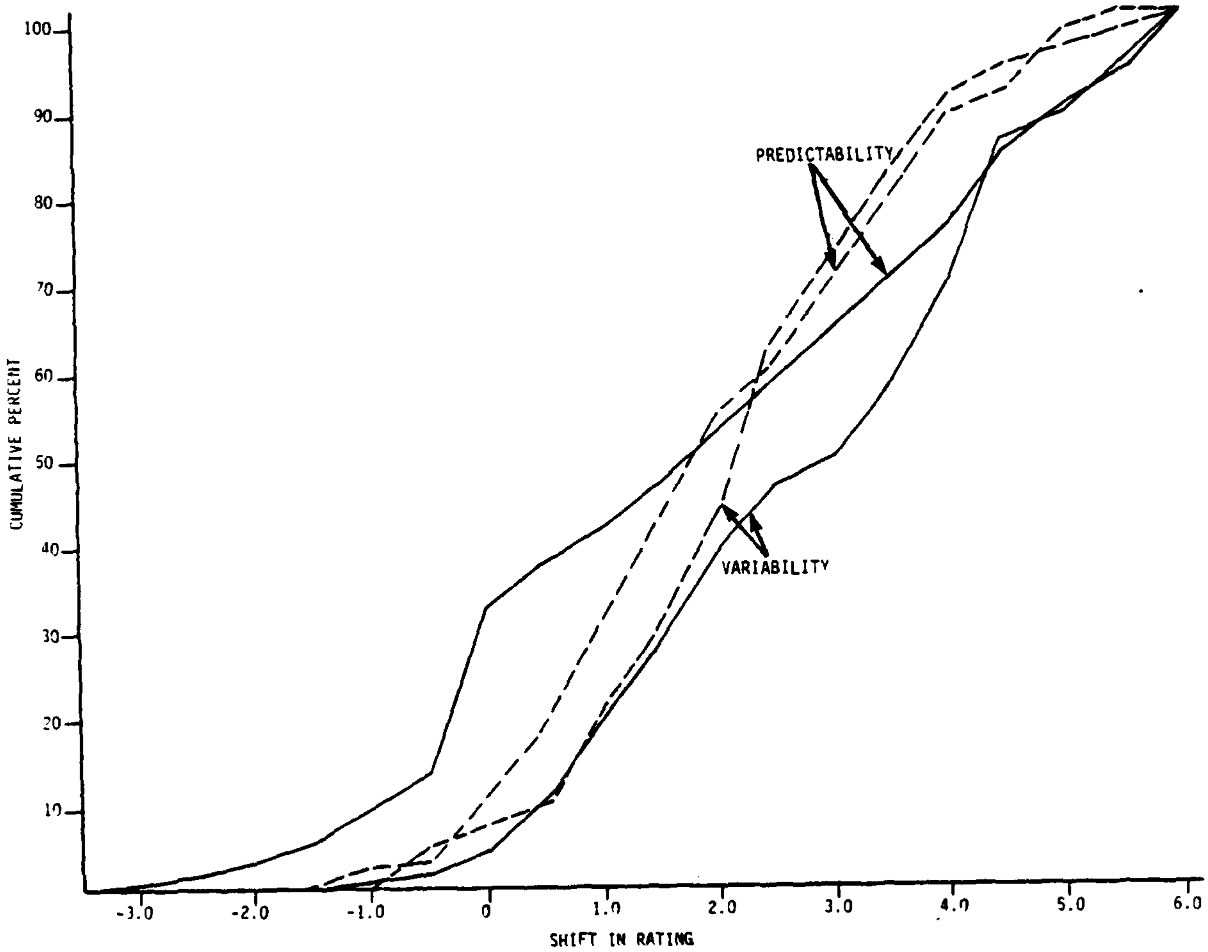


FIGURE 3-10 (CONT.)

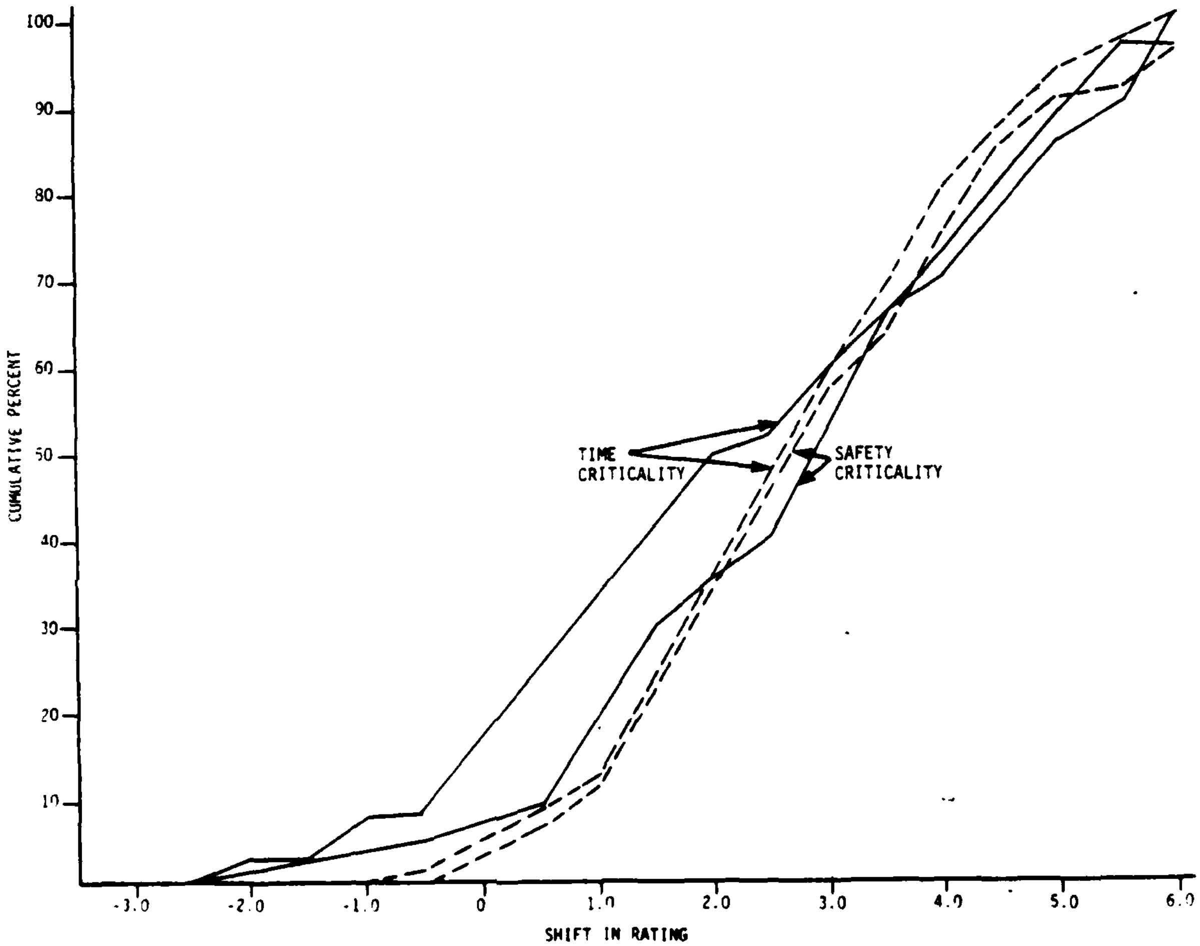


FIGURE 3-10 (CONT.)

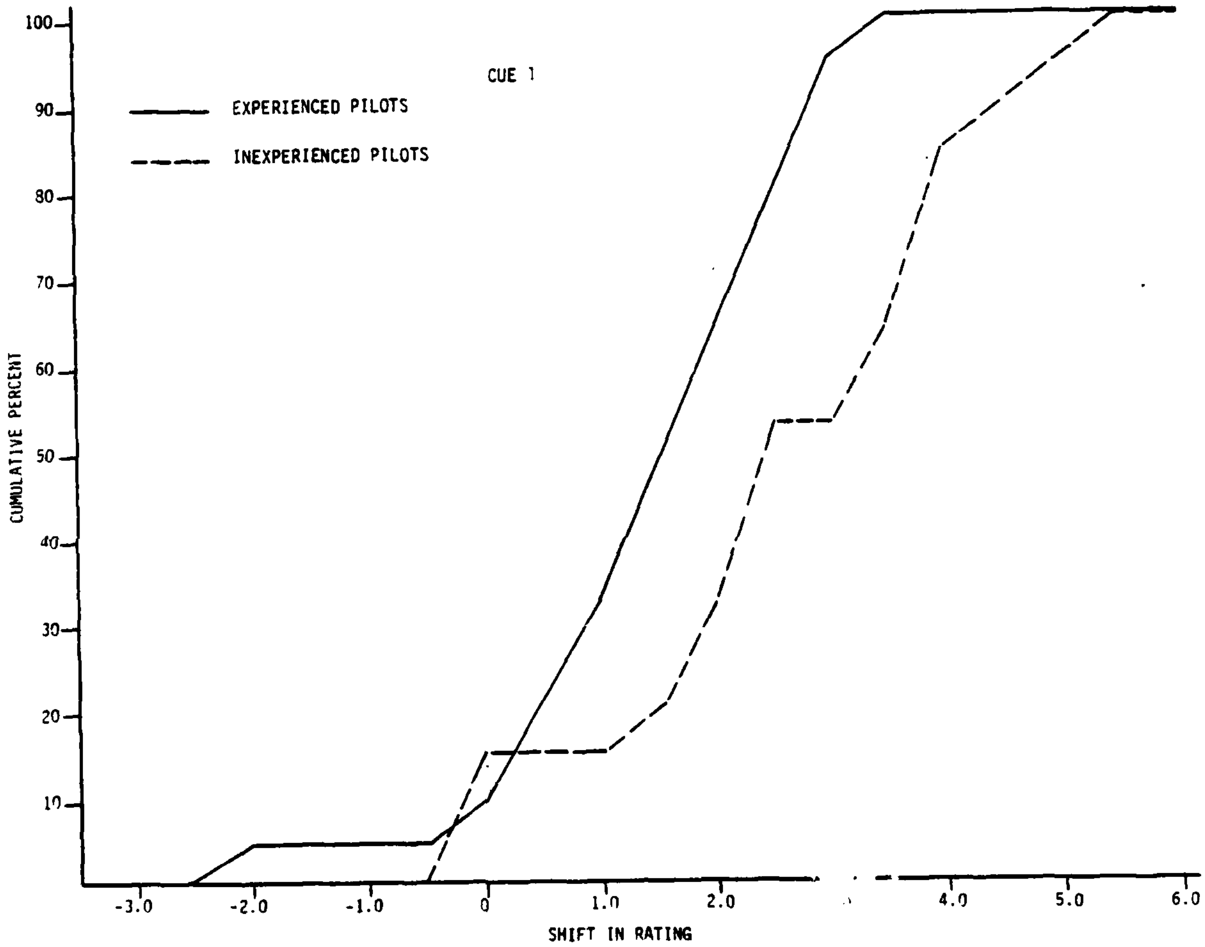


FIGURE 3-11.
SHIFTS IN RATINGS OF SALIENCE BY CUE

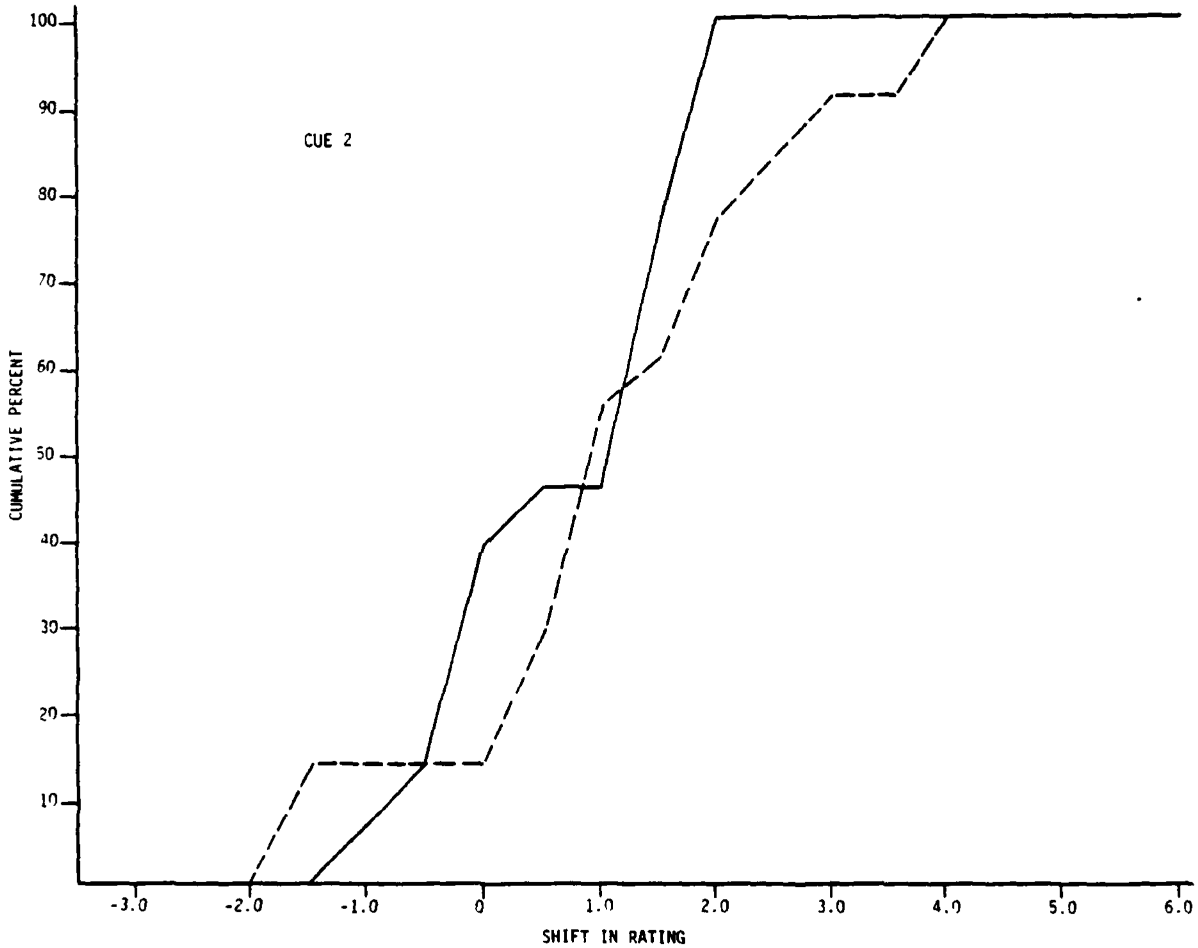


FIGURE 3-11. (CONT.)

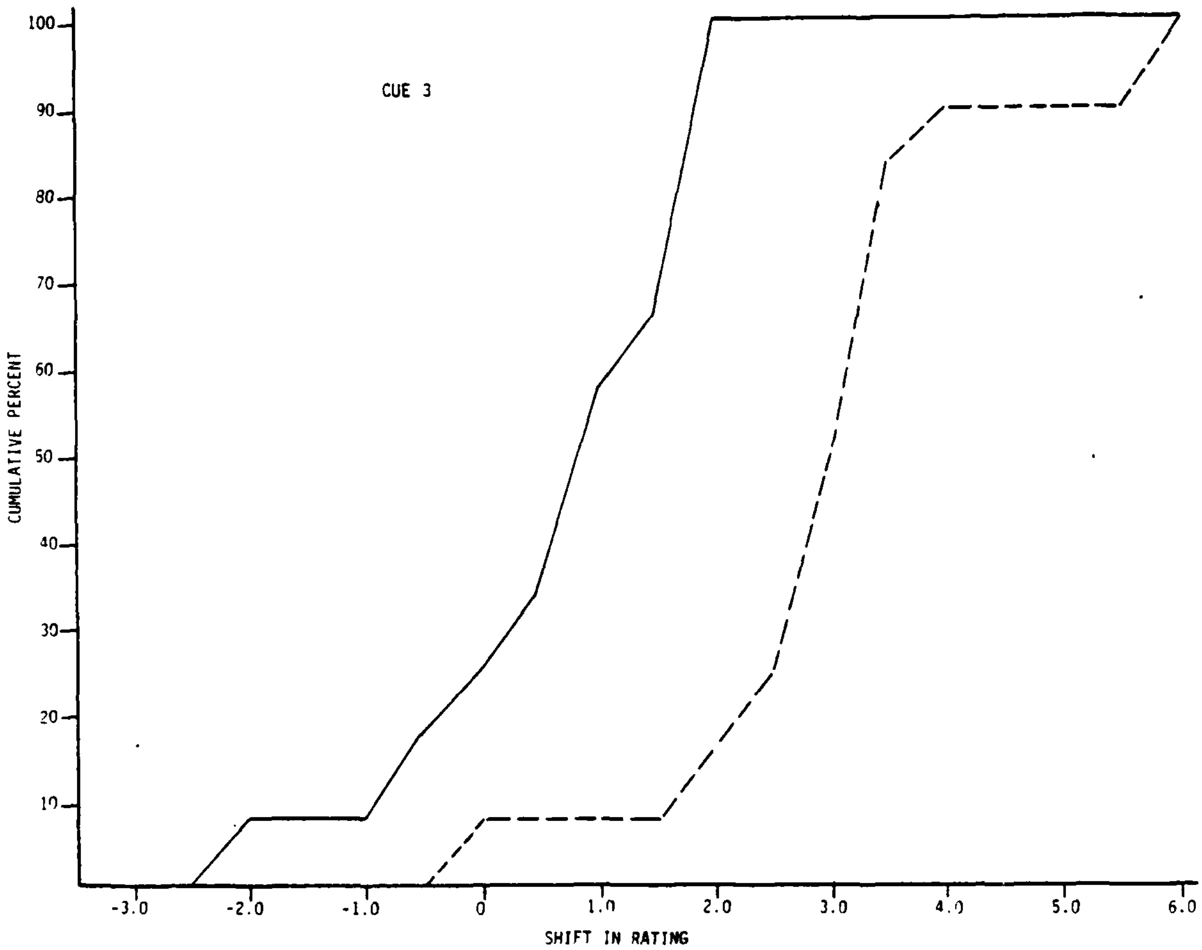


FIGURE 3-11. (CONT.)

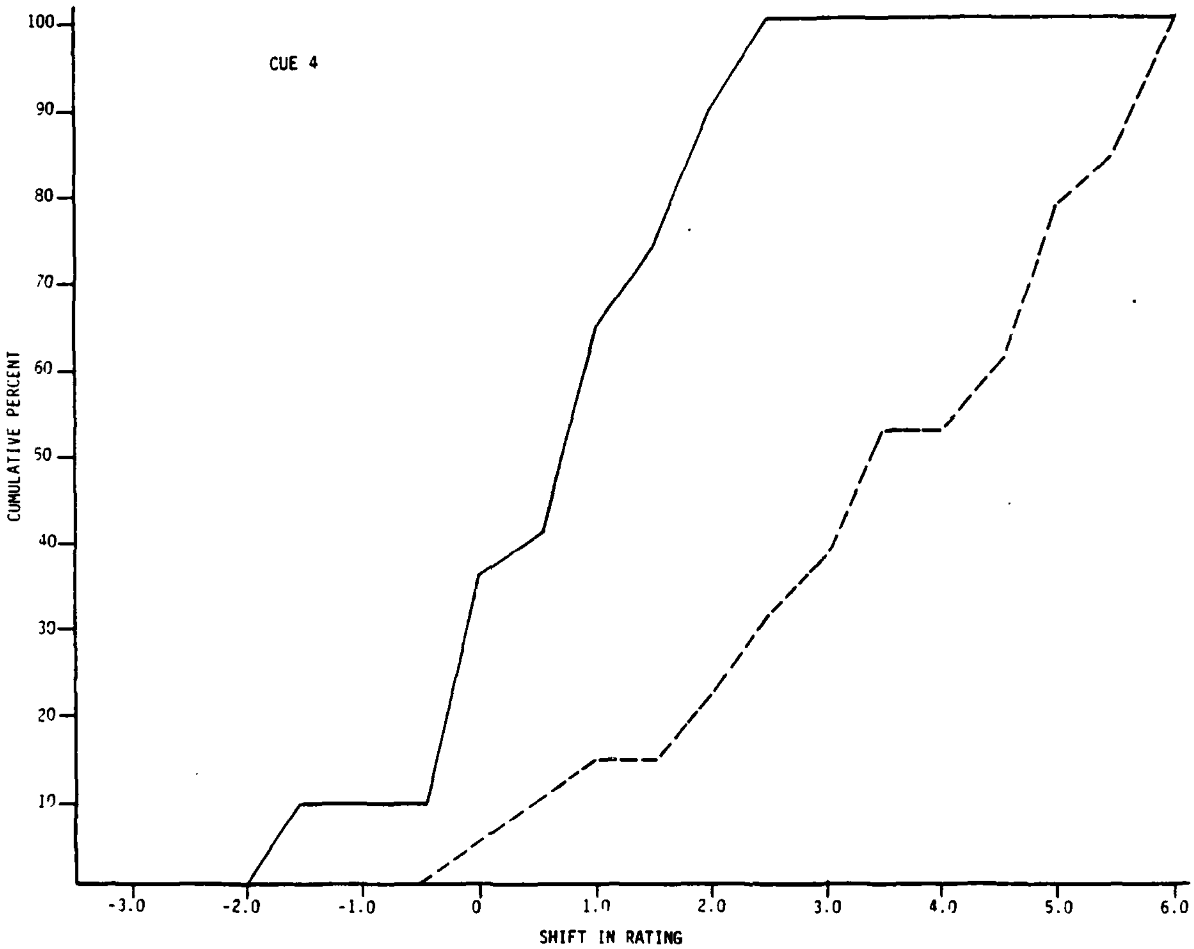


FIGURE 3-11 (CONT.)

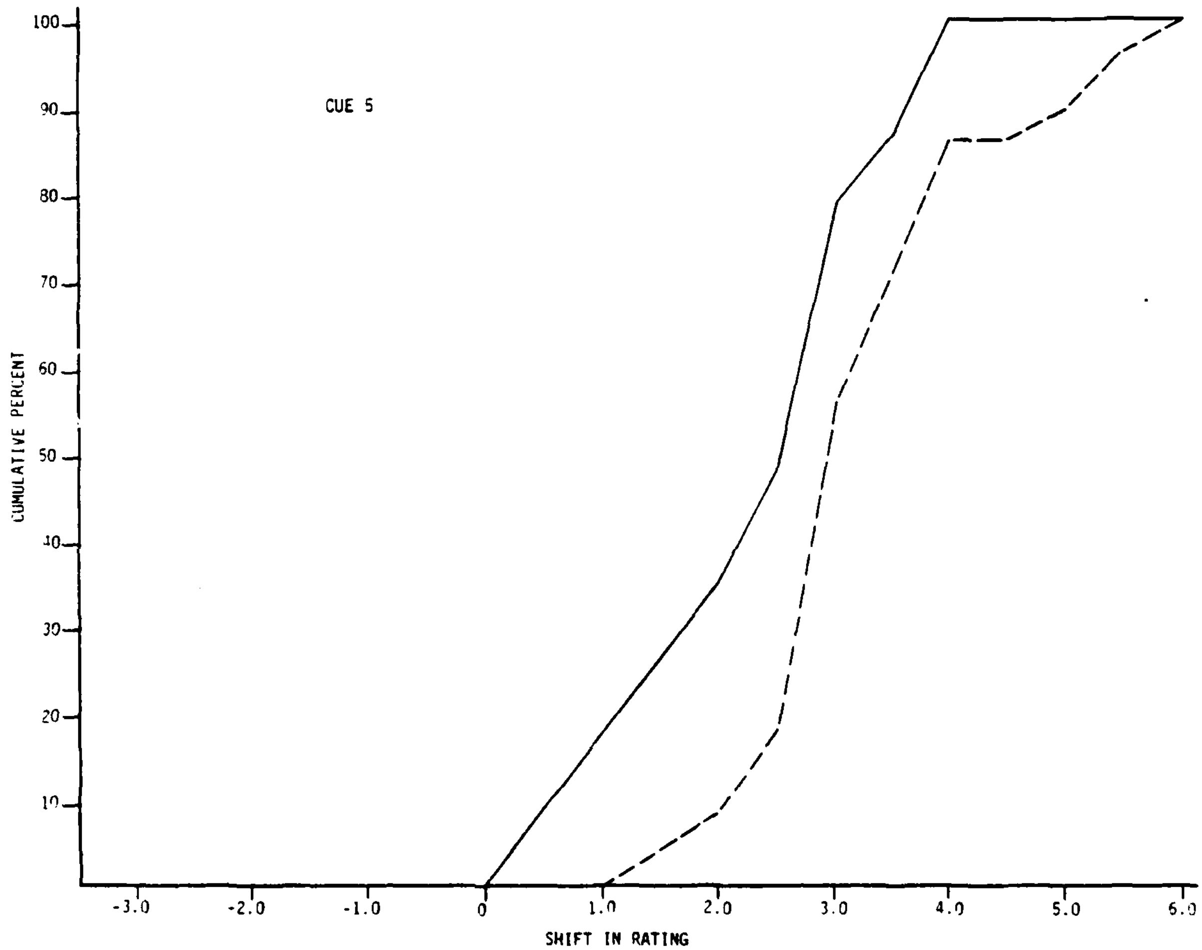


FIGURE 3-11. (CONT.)

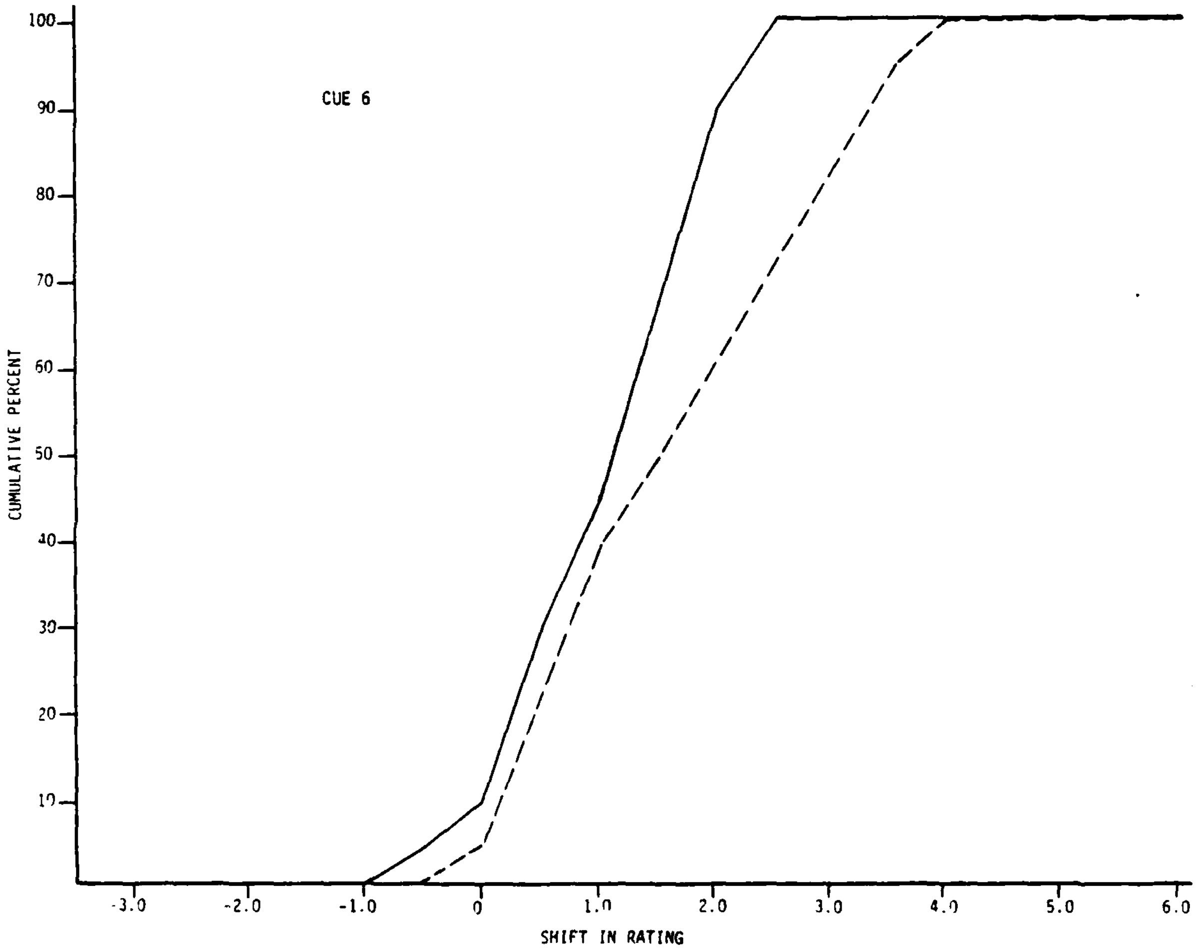


FIGURE 3-11. (CONT.)

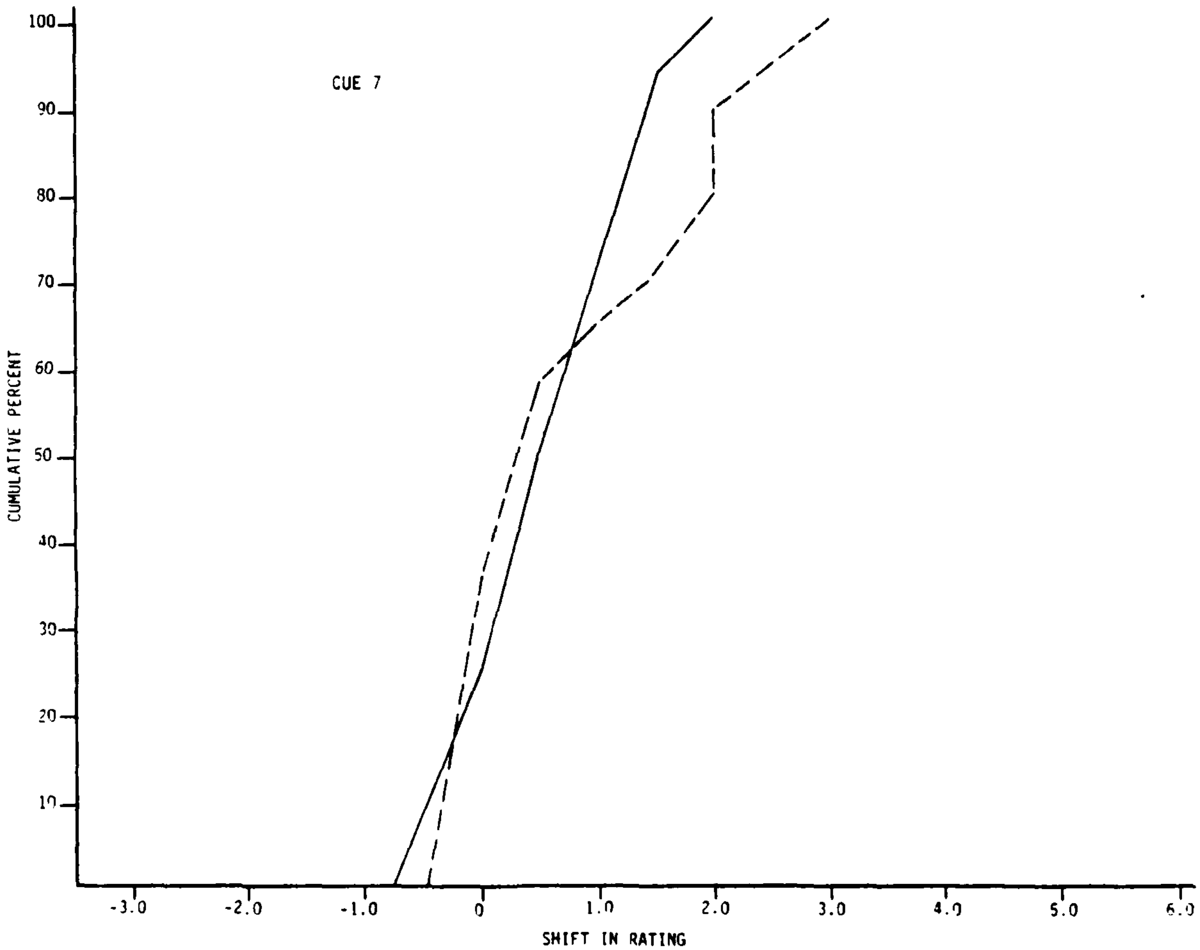


FIGURE 3-11 (CONT.)

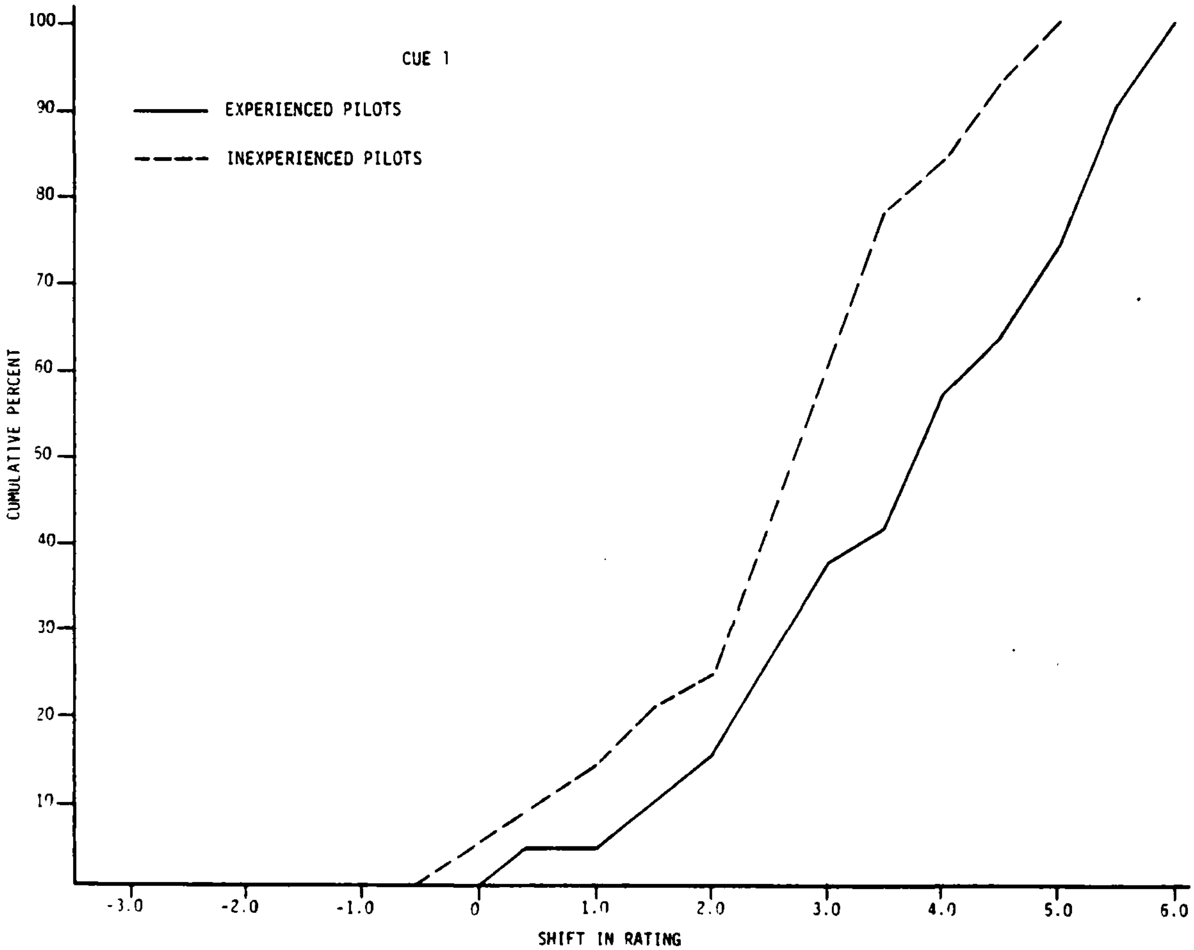


FIGURE 3-12.
SHIFTS IN RATINGS OF DIAGNOSTICITY BY CUE

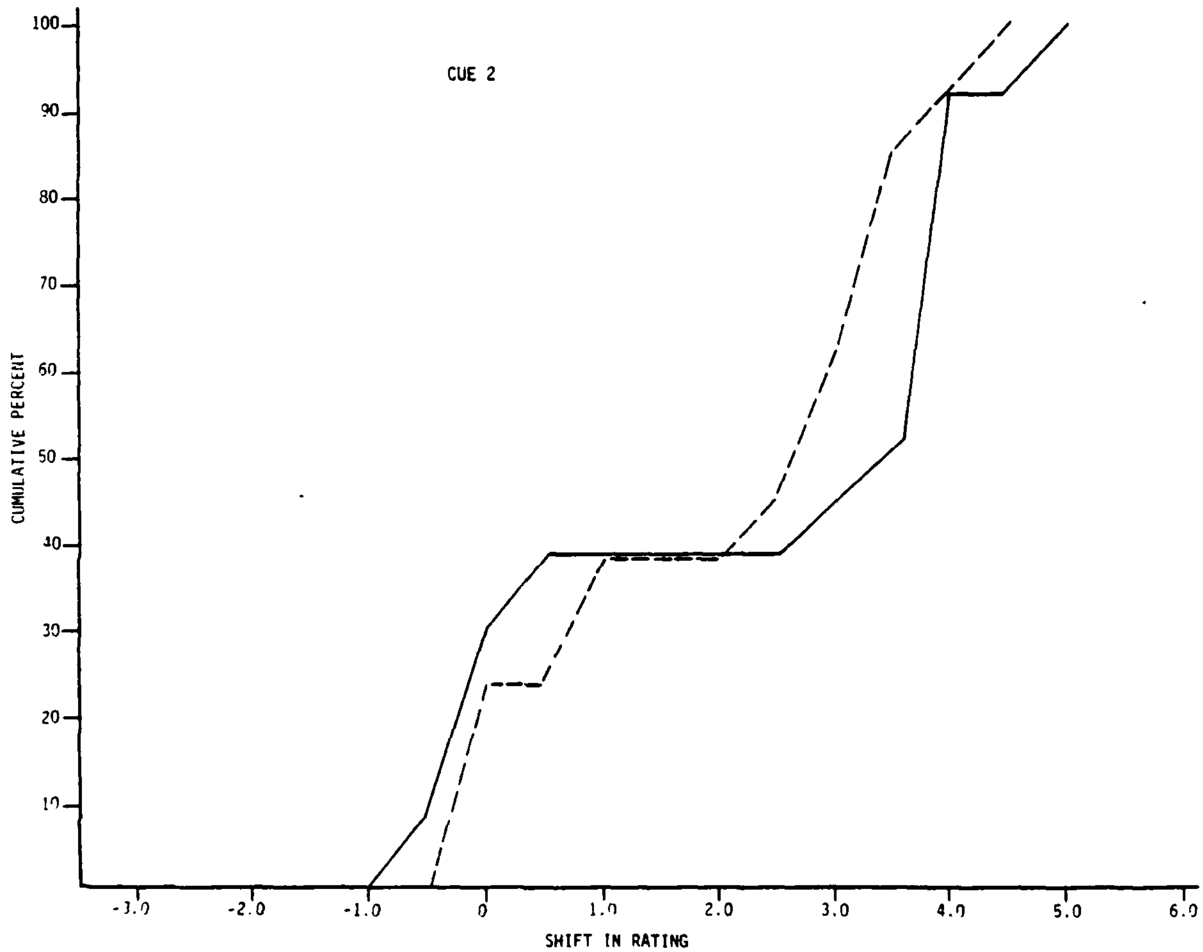


FIGURE 3-12. (CONT.)

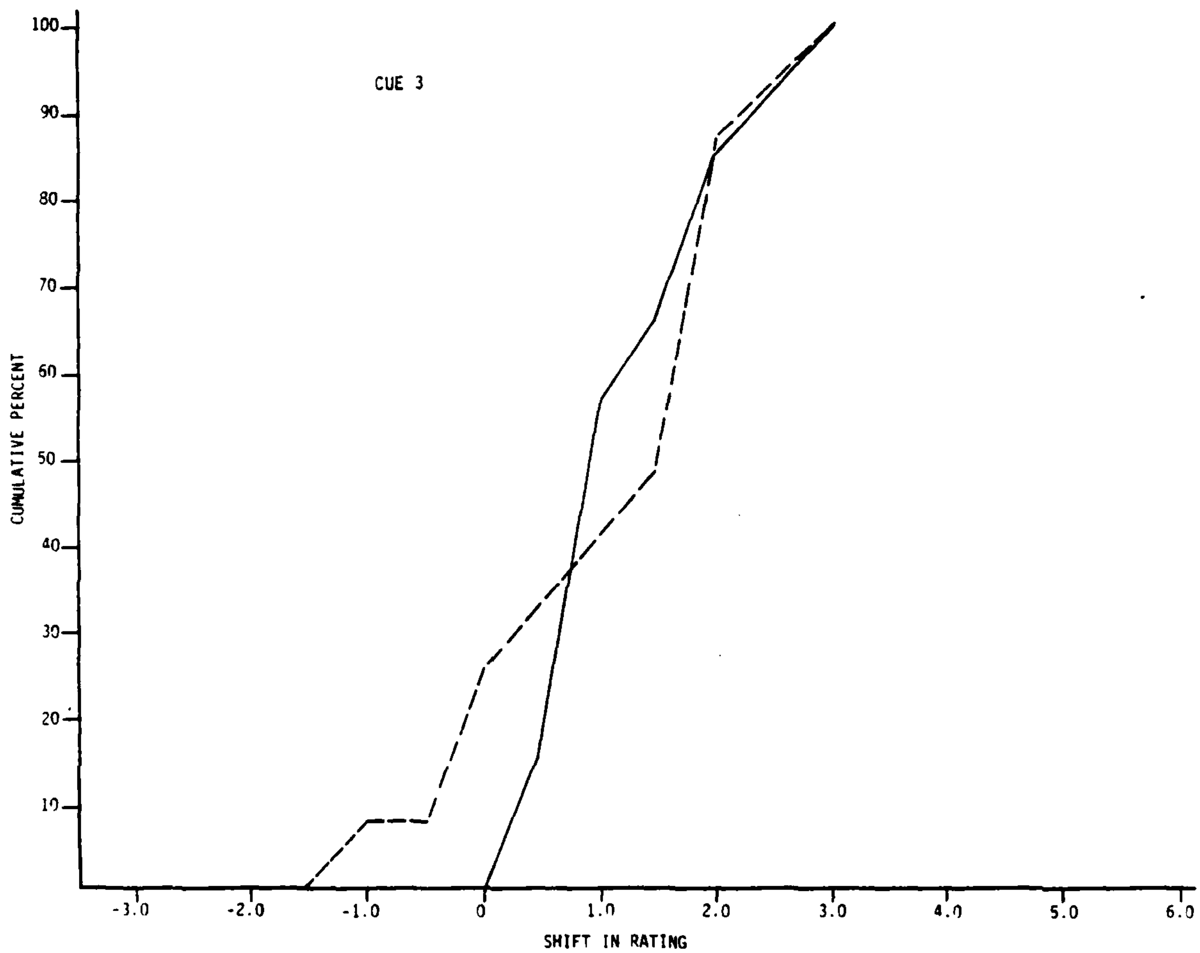


FIGURE 3-12. (CONT.)

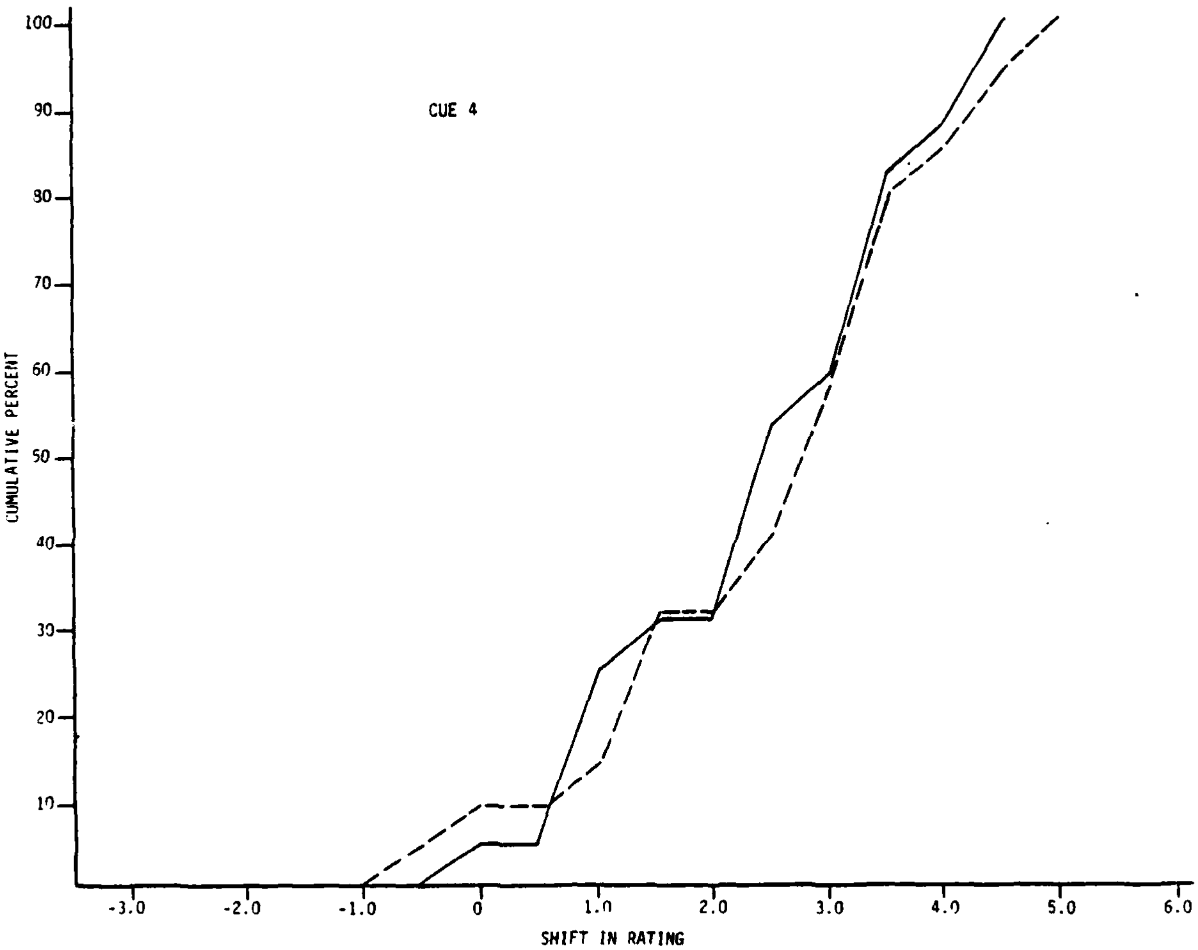


FIGURE 3-12. (CONT.)

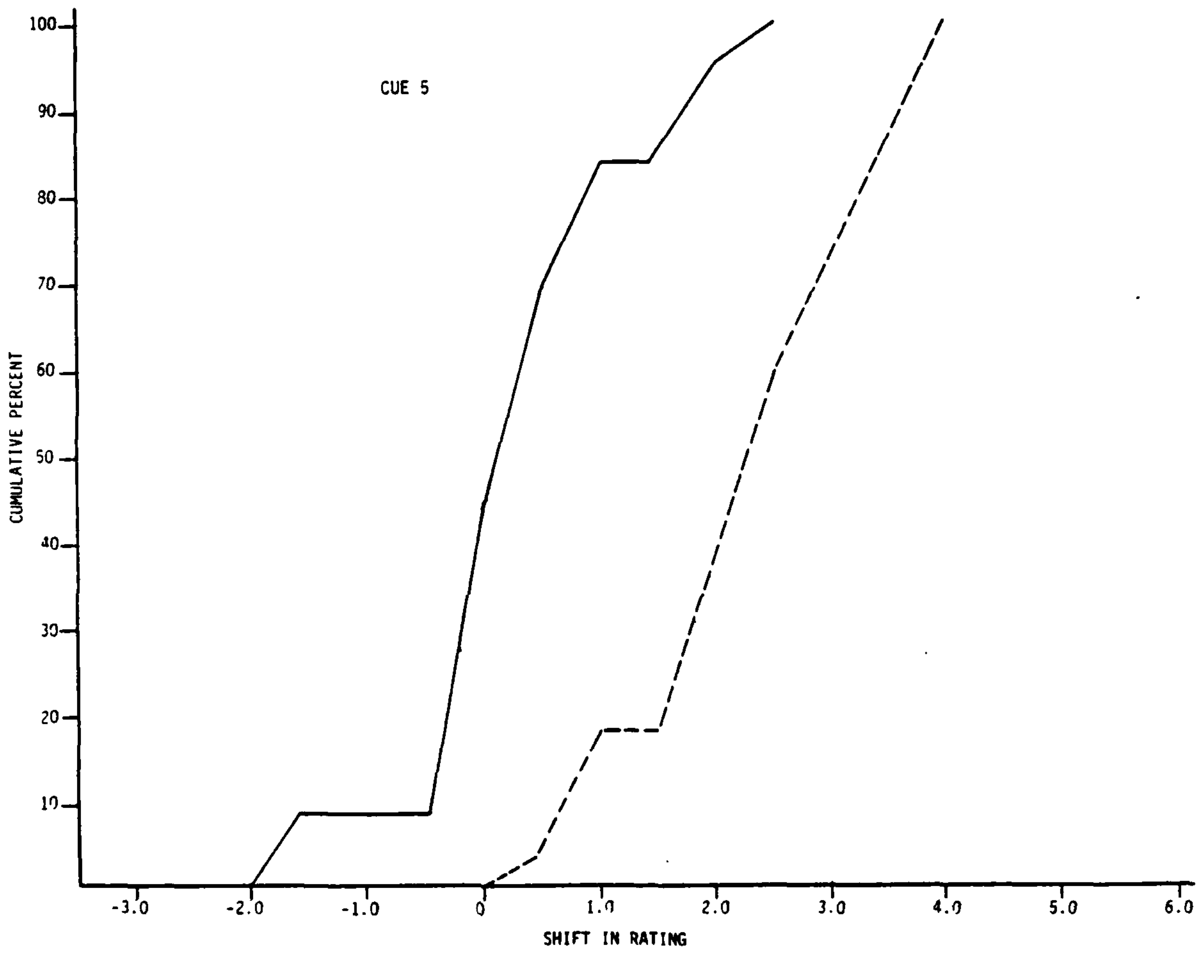


FIGURE 3-12. (CONT.)

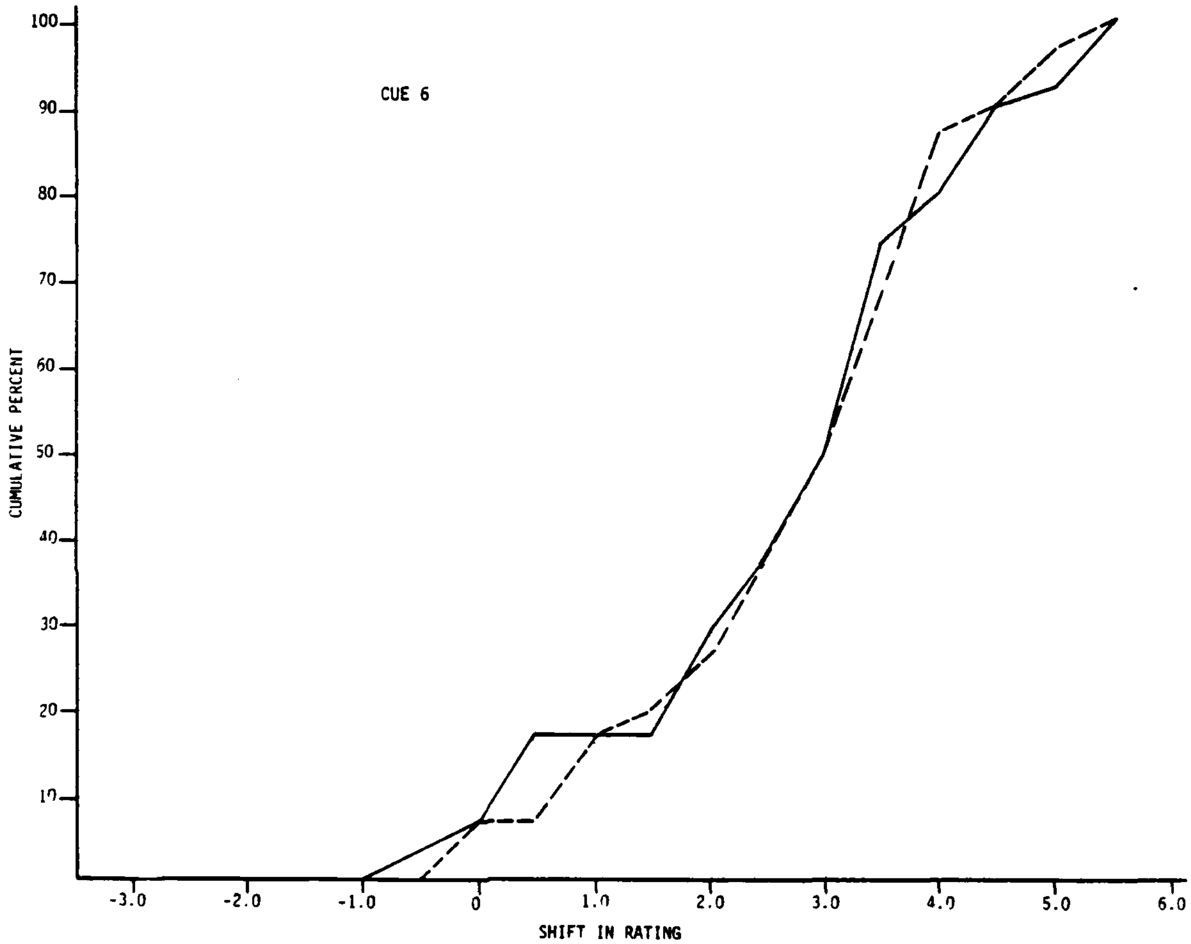


FIGURE 3-12. (CONT.)

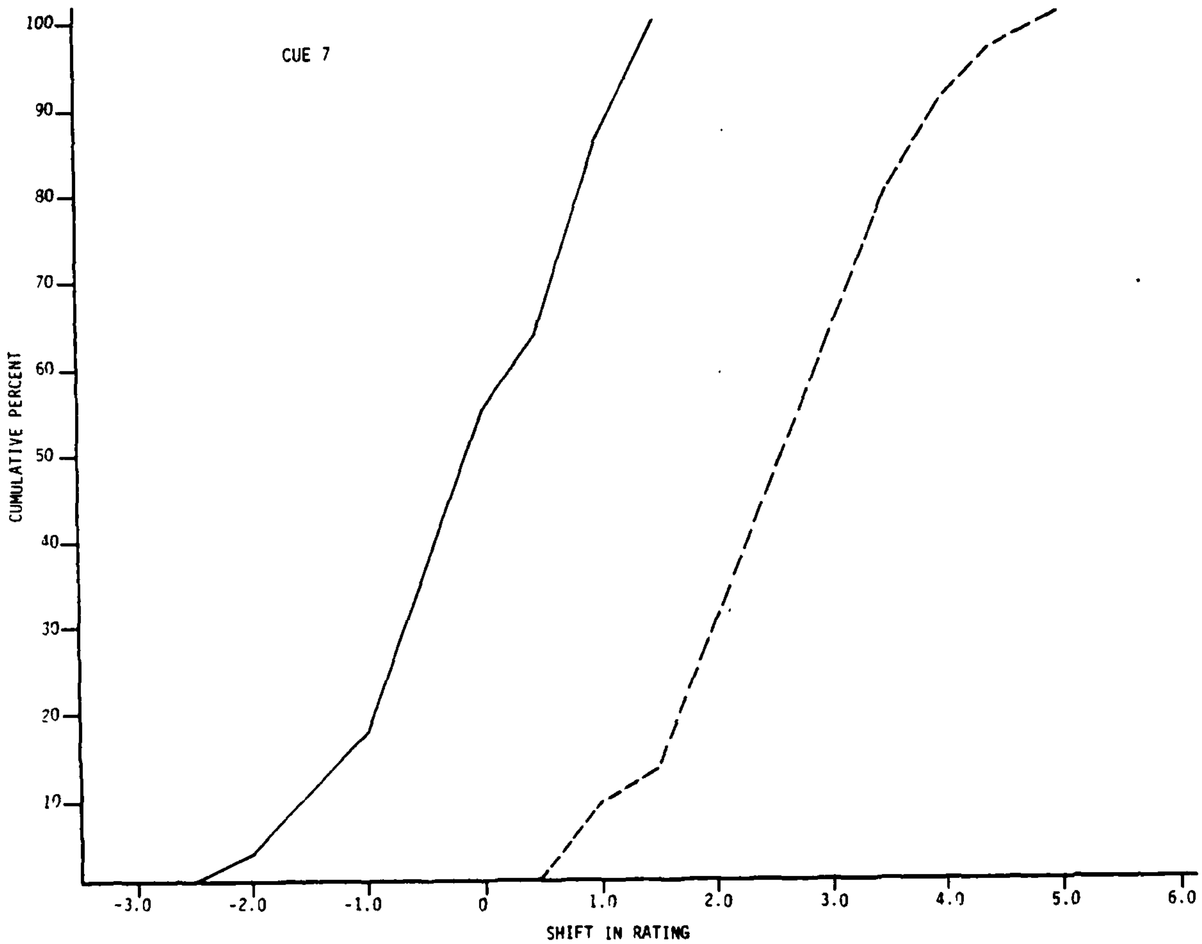


FIGURE 3-12. (CONT.)

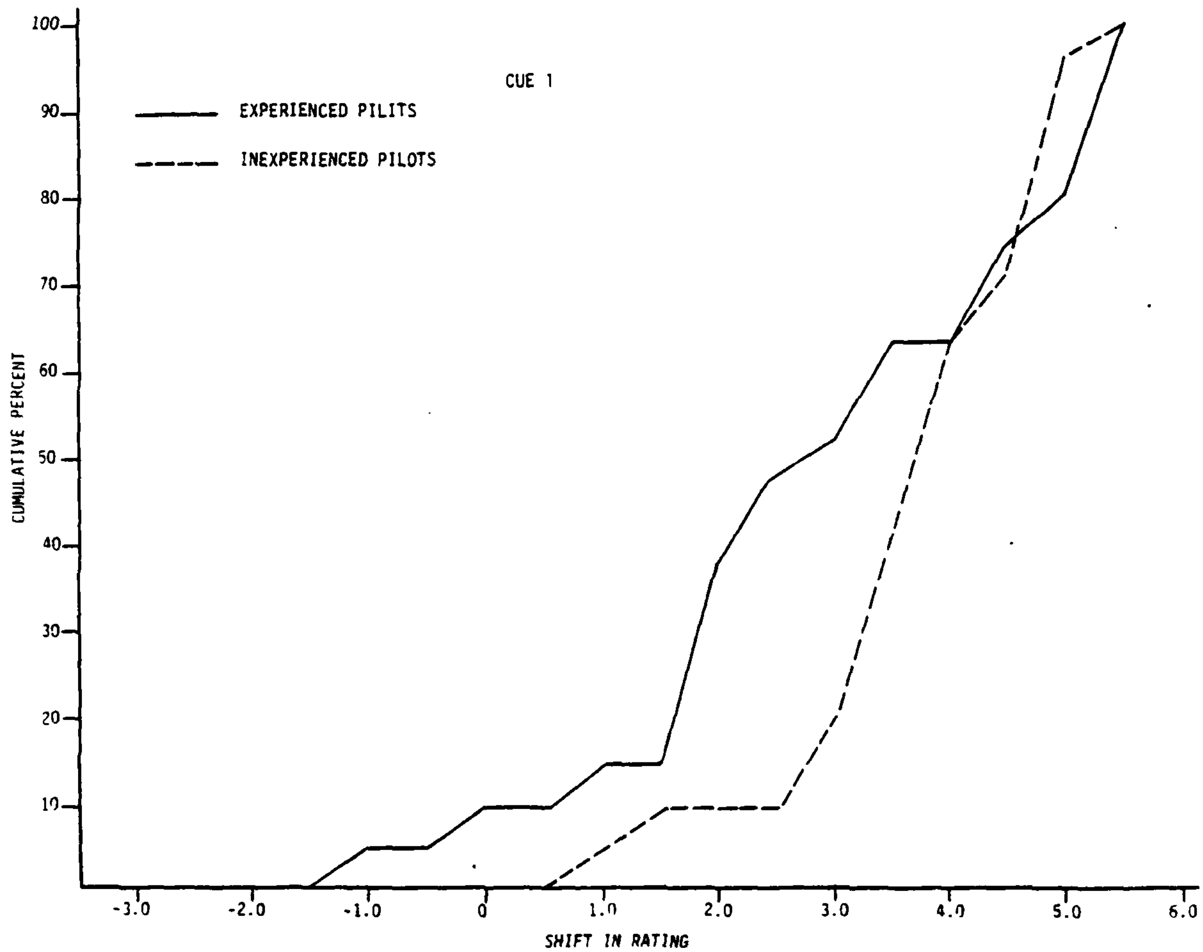


FIGURE 3-13.
SHIFTS IN RATINGS OF PREDICTABILITY BY CUE

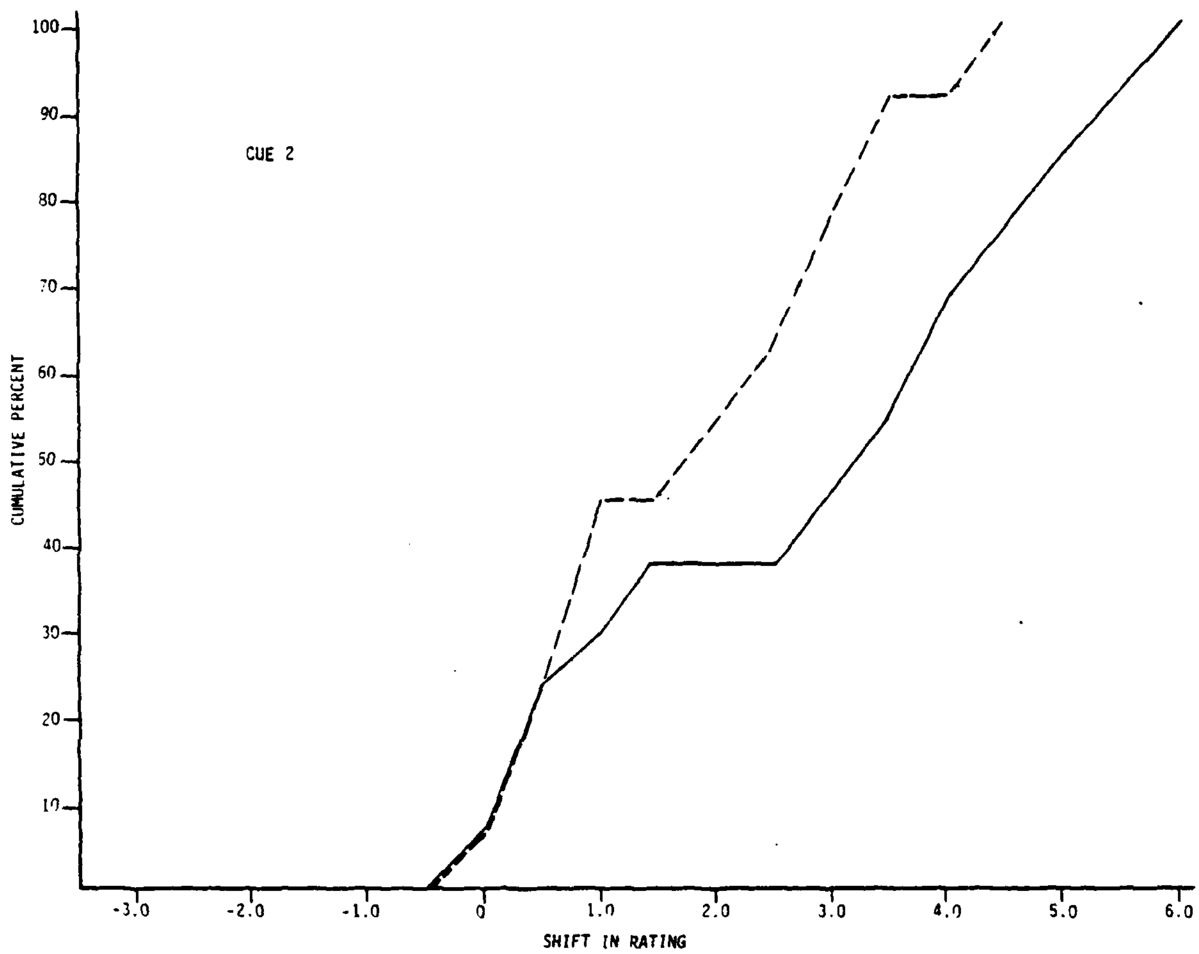


FIGURE 3-13. (CONT.)

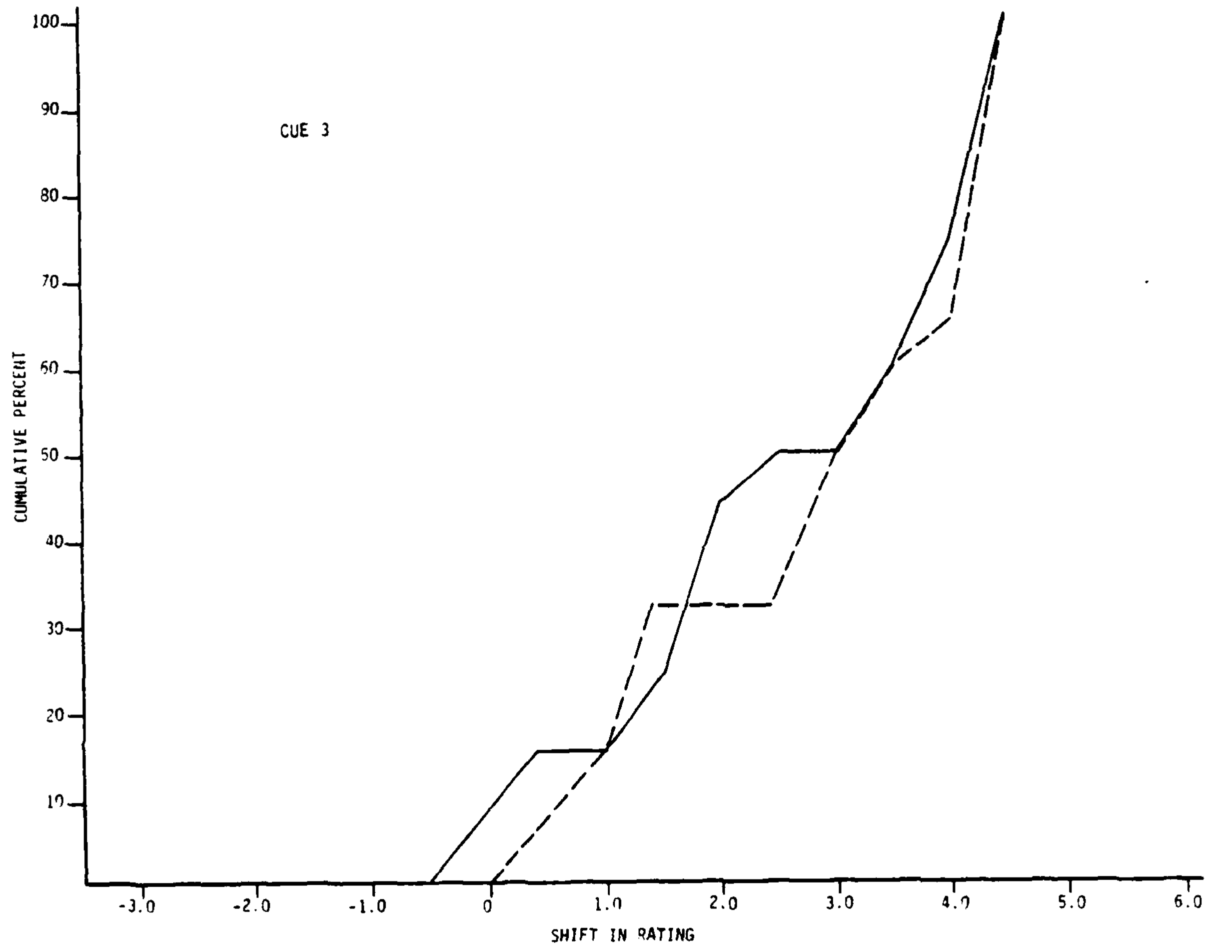


FIGURE 3-13. (CONT.)

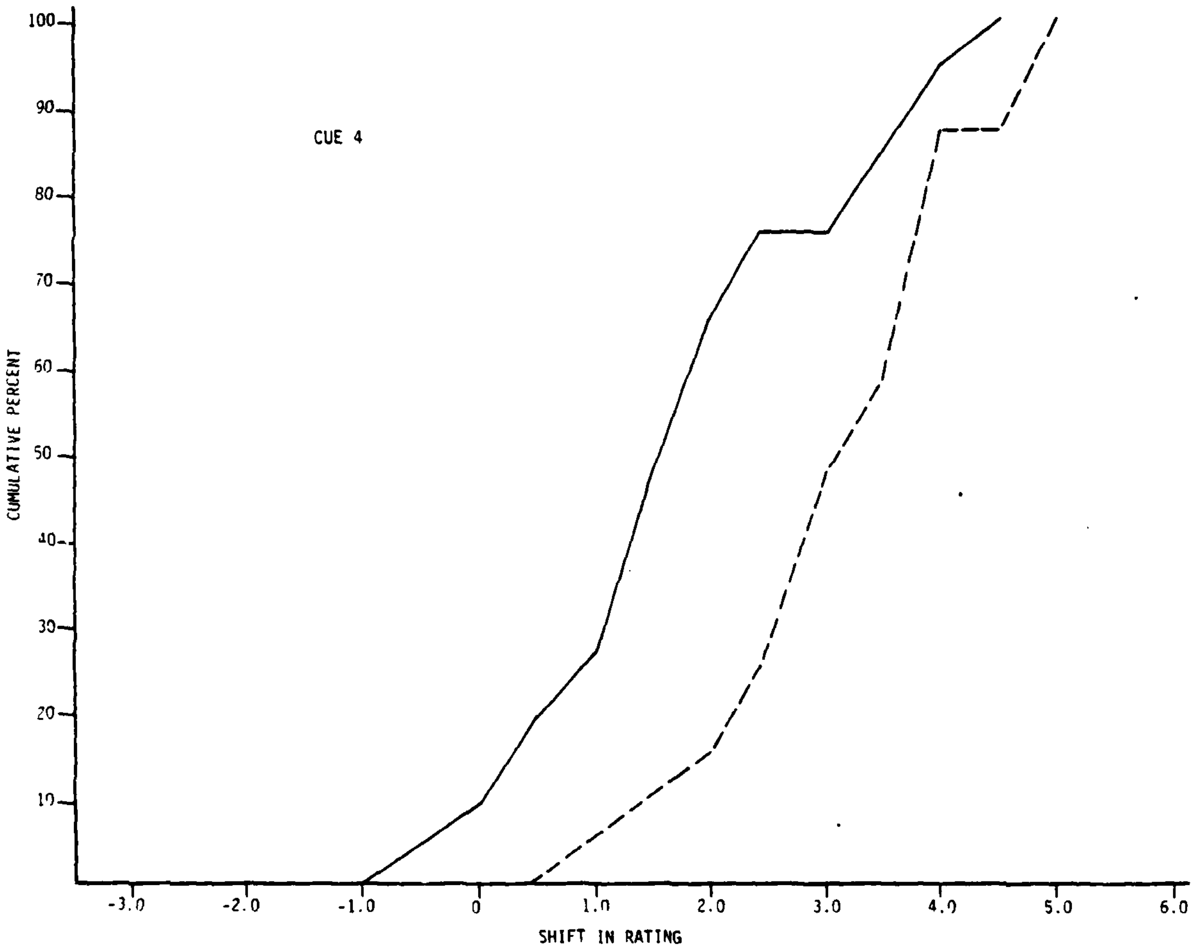


FIGURE 3-13. (CONT.)

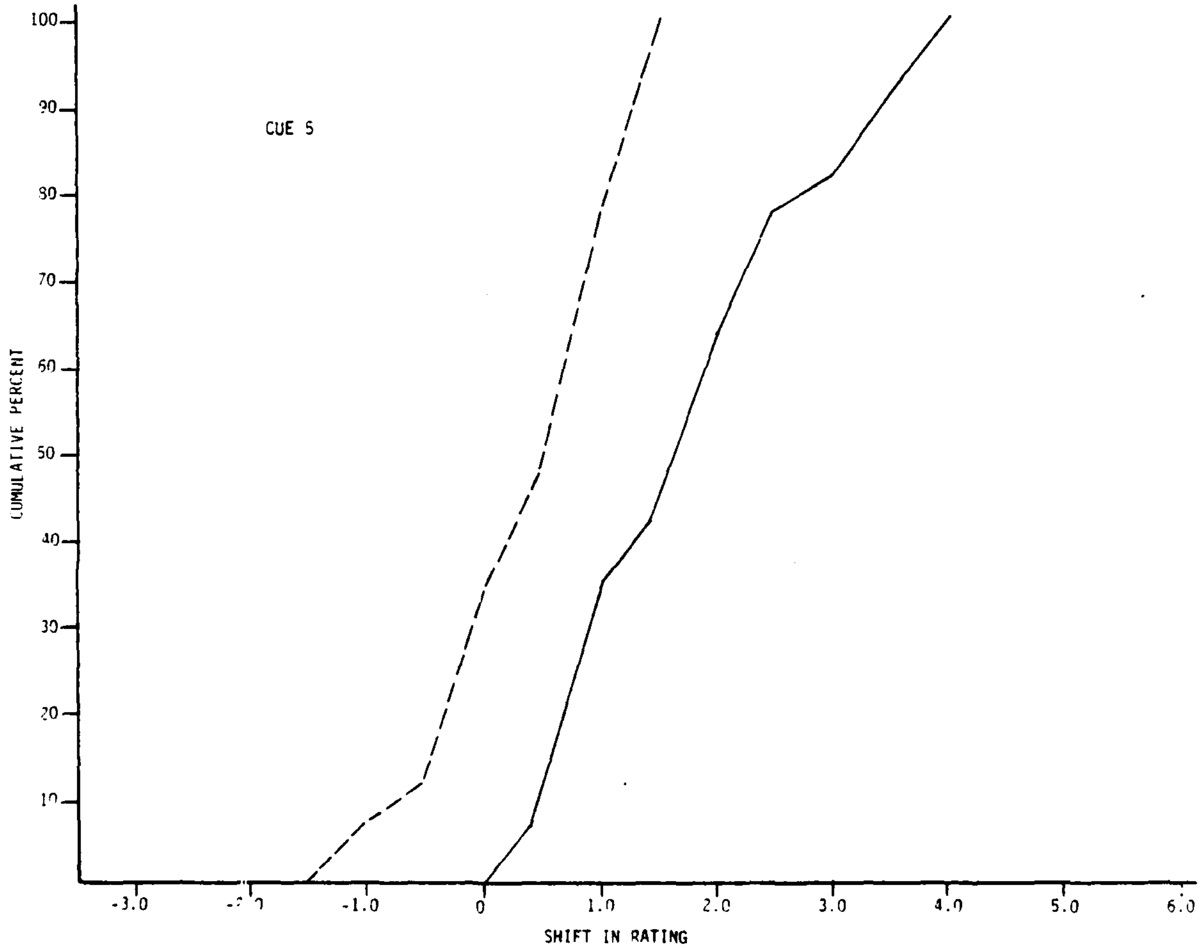


FIGURE 3-13. (CONT.)

AD-A100 140 PERCEPTRONICS INC WOODLAND HILLS CALIF F/G 5/9
AIRCREW EMERGENCY DECISIONS: ANALYSIS OF CUE FACTORS FOR USE IN--ETC(U)
DEC 80 L F LUCACCINI, R HOPF-WEICHEL F49620-78-C-0067
UNCLASSIFIED PFTR-1065-80-12 AFOSR-TR-81-0475 NL

2 OF 2
AD
A100140

END
DATE
FILMED
7 81
DTIC

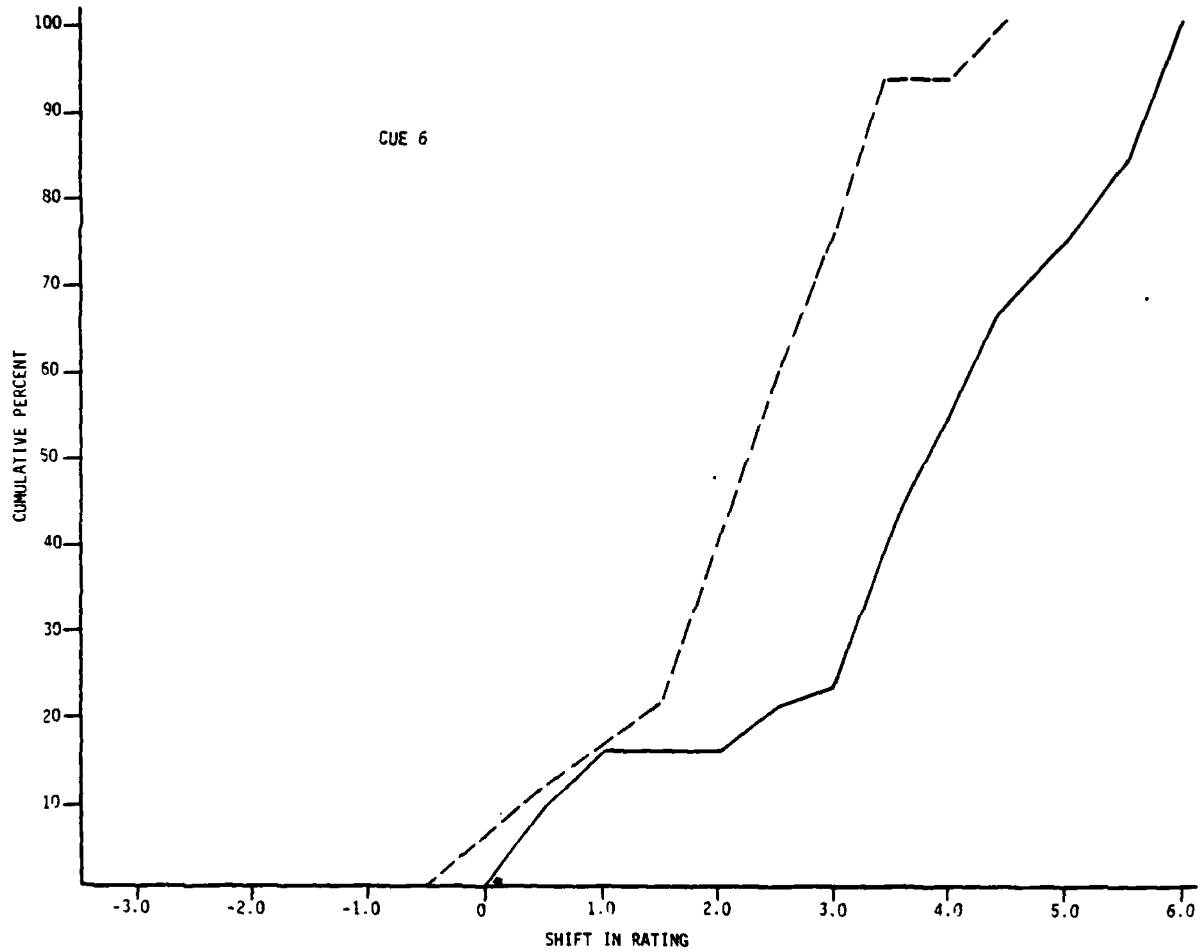


FIGURE 3-13. (CONT.)

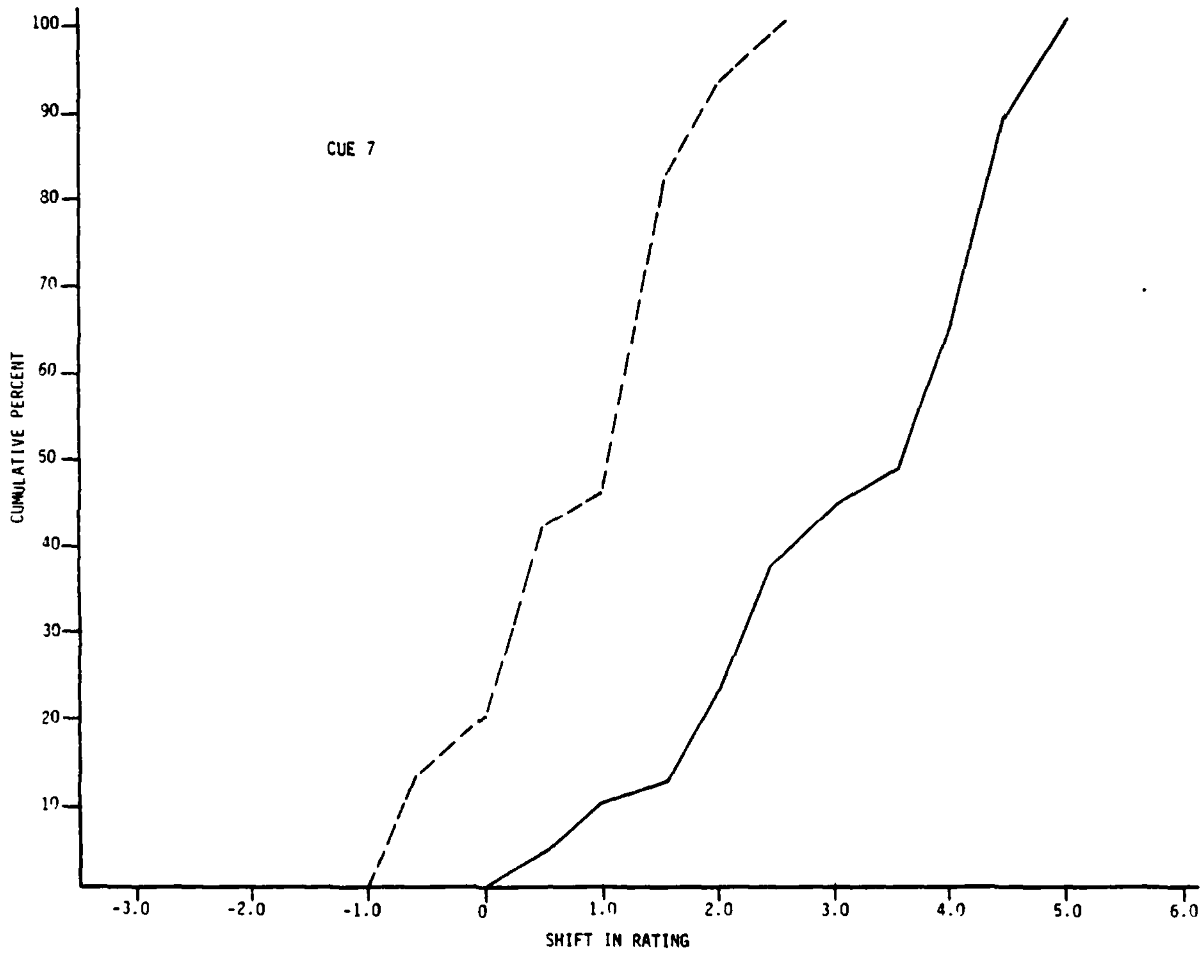


FIGURE 3-13. (CONT.)

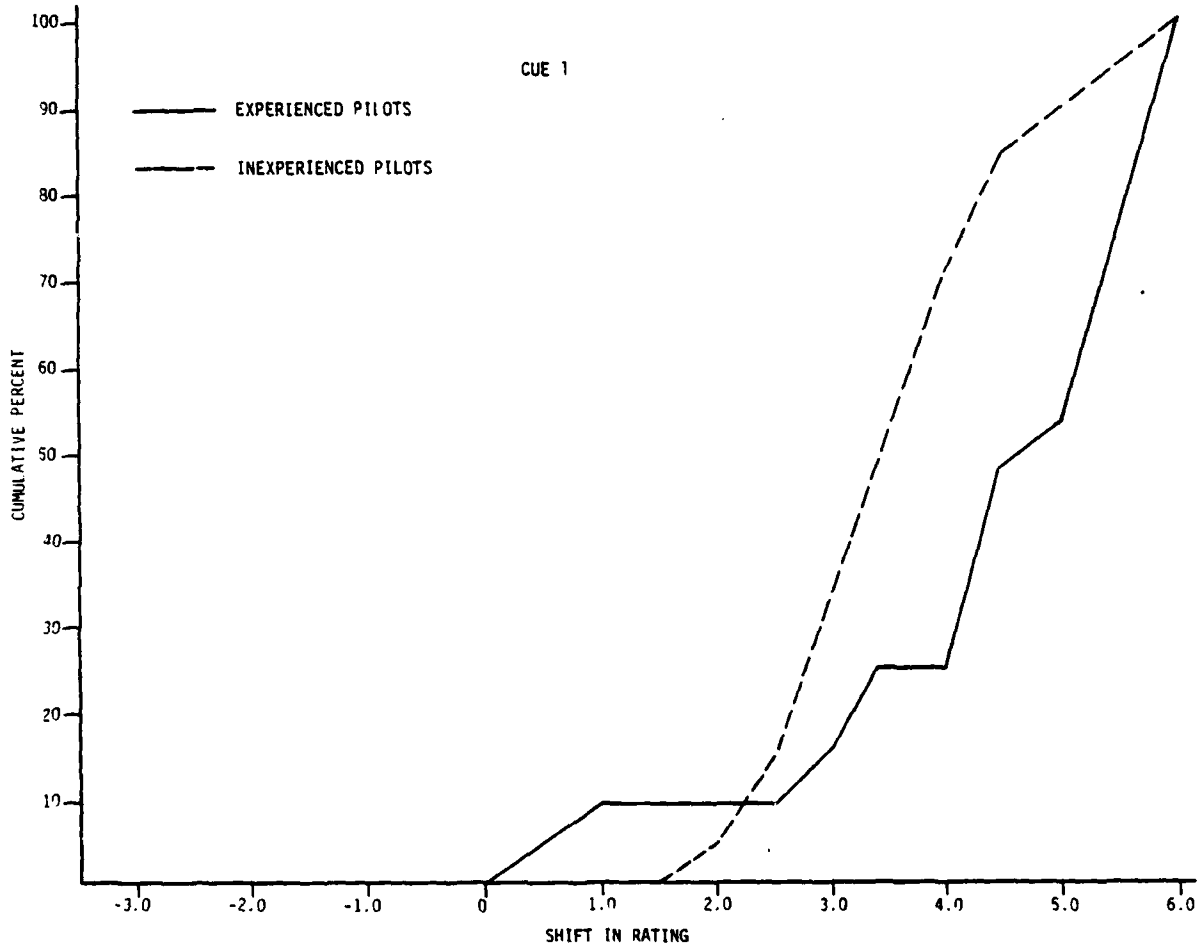


FIGURE 3-14.
SHIFTS IN RATINGS OF VARIABILITY BY CUE

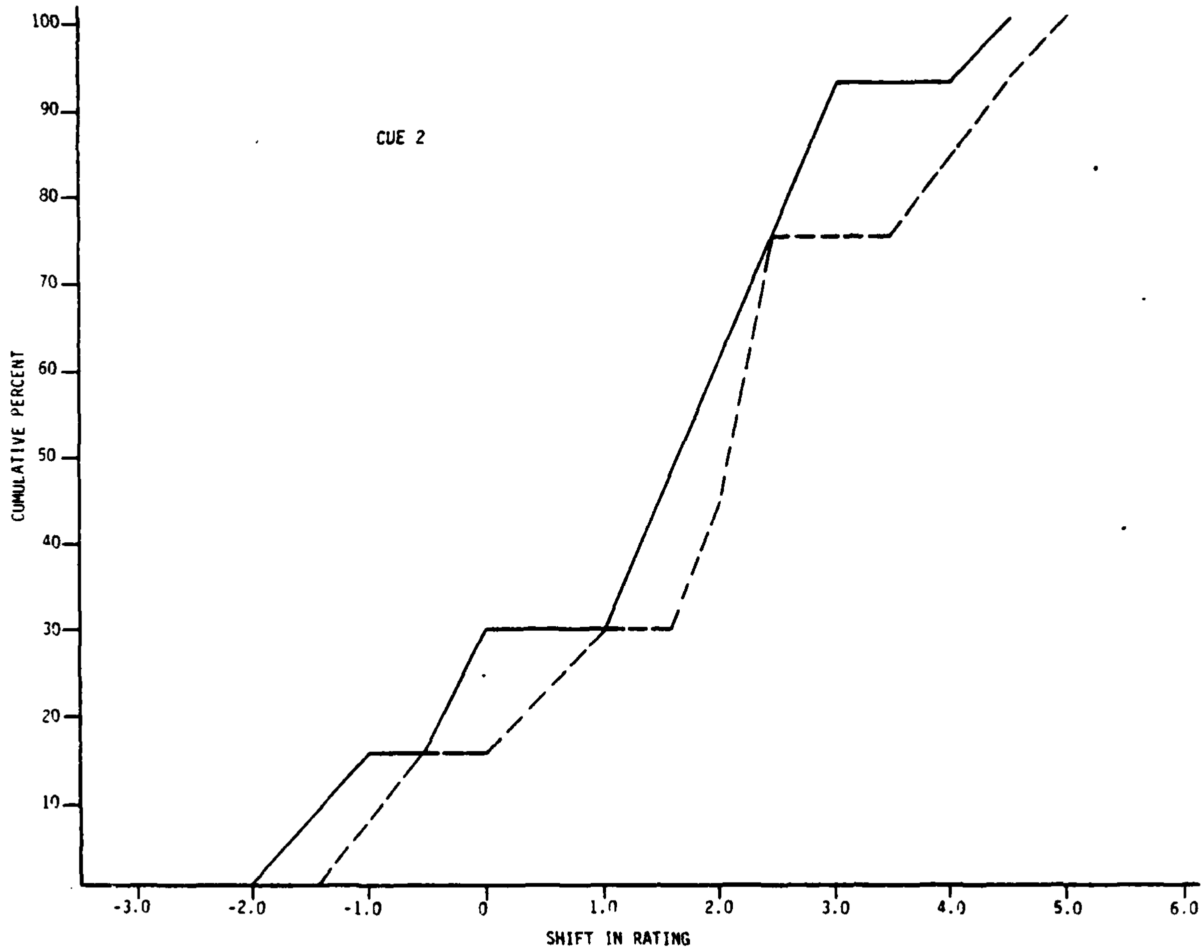


FIGURE 3-14. (CONT.)

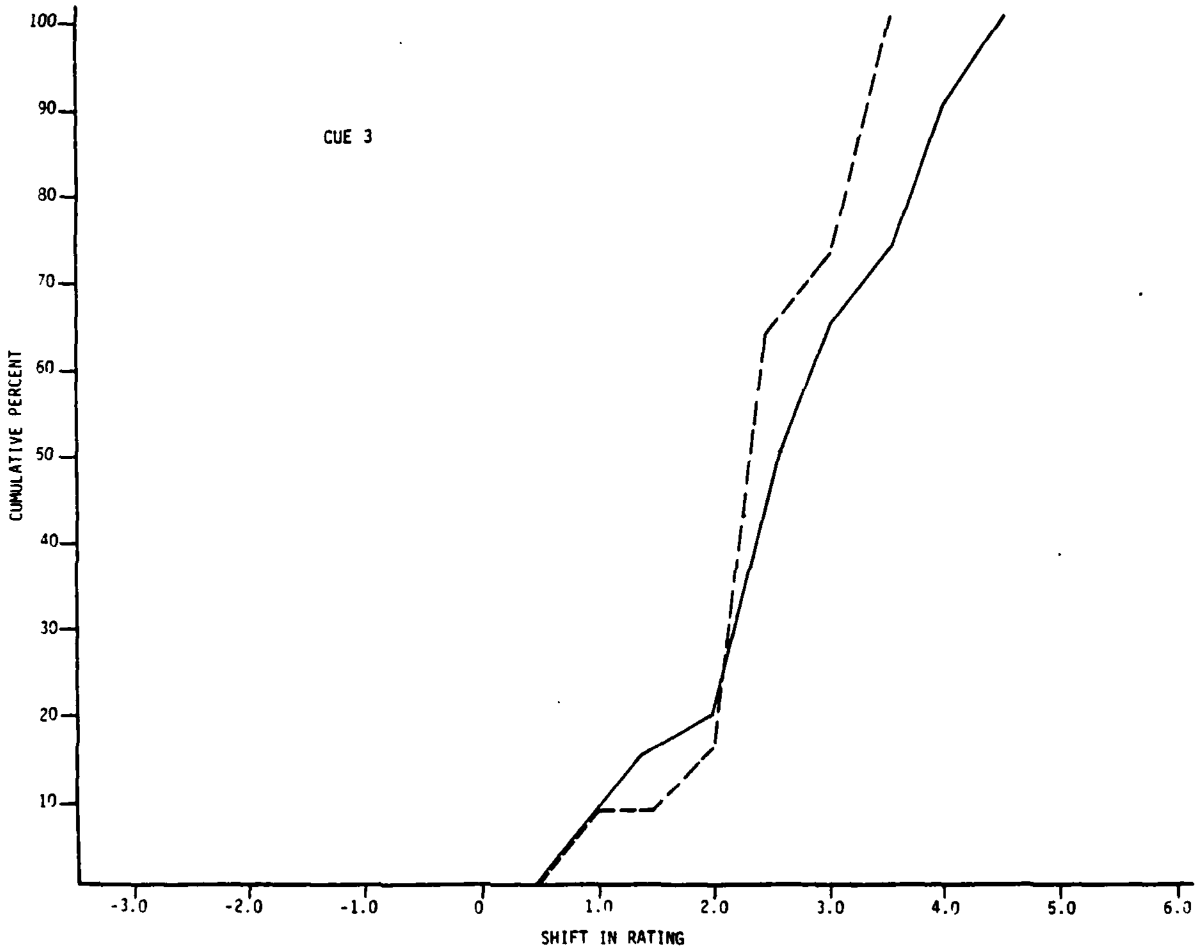


FIGURE 3-14. (CONT.)

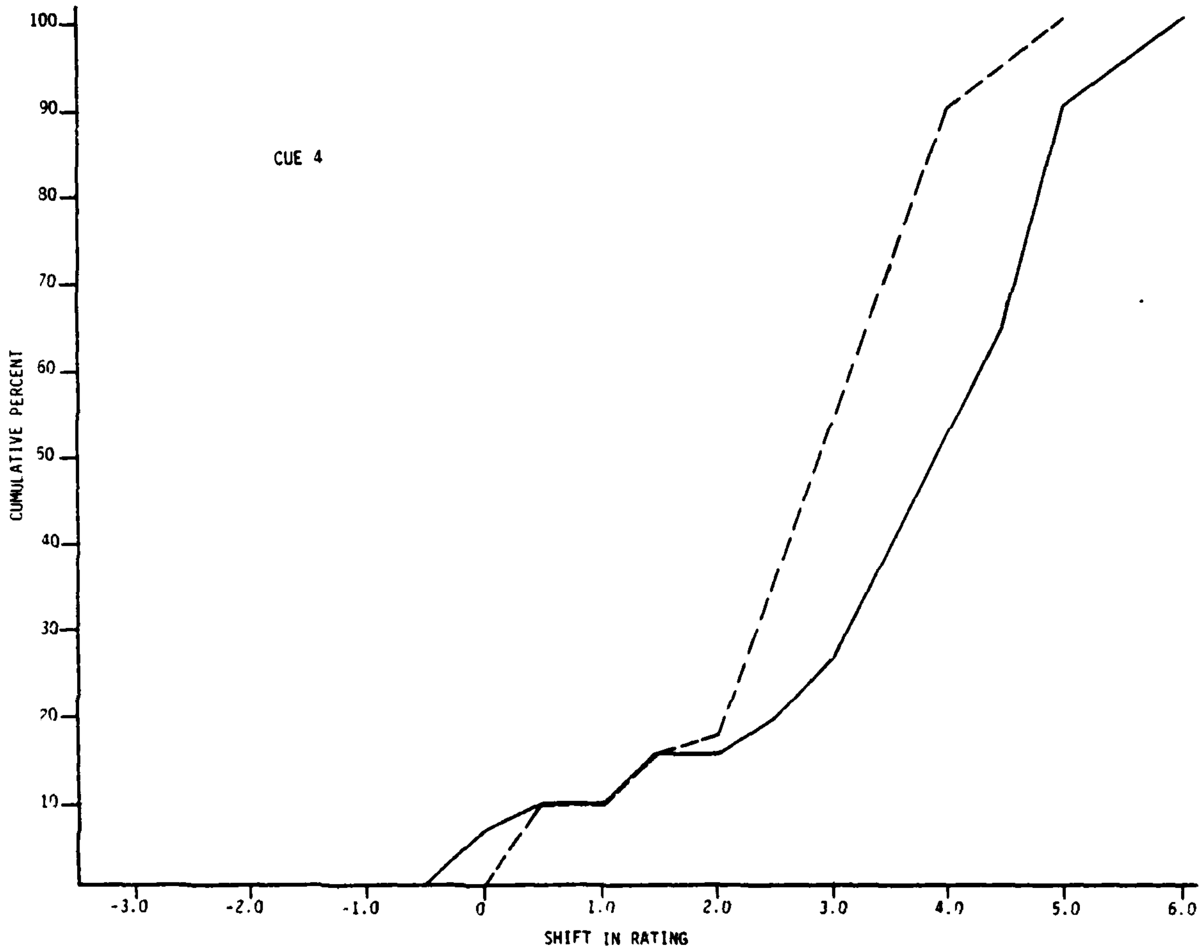


FIGURE 3-14. (CONT.)

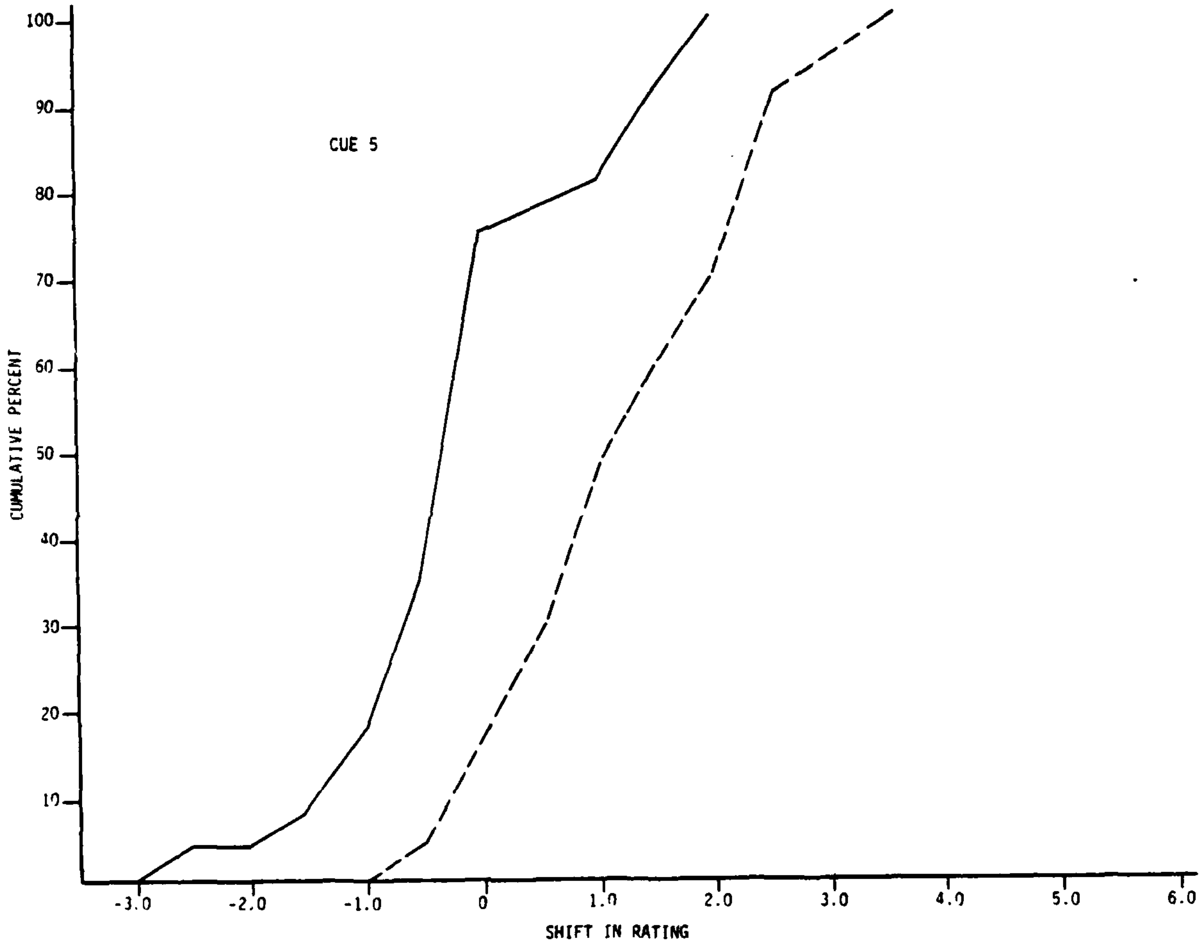


FIGURE 3-14. (CONT.)

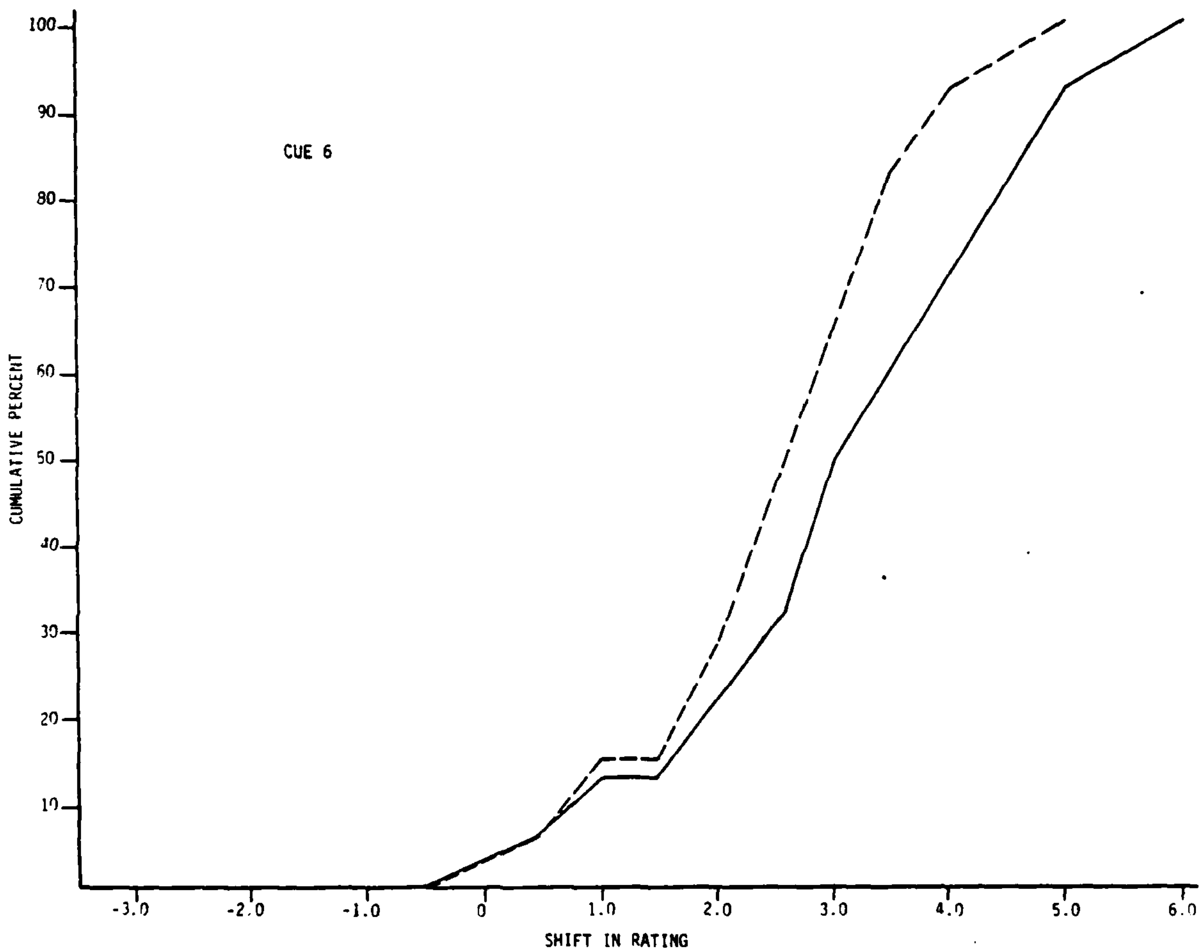


FIGURE 3-14. (CONT.)

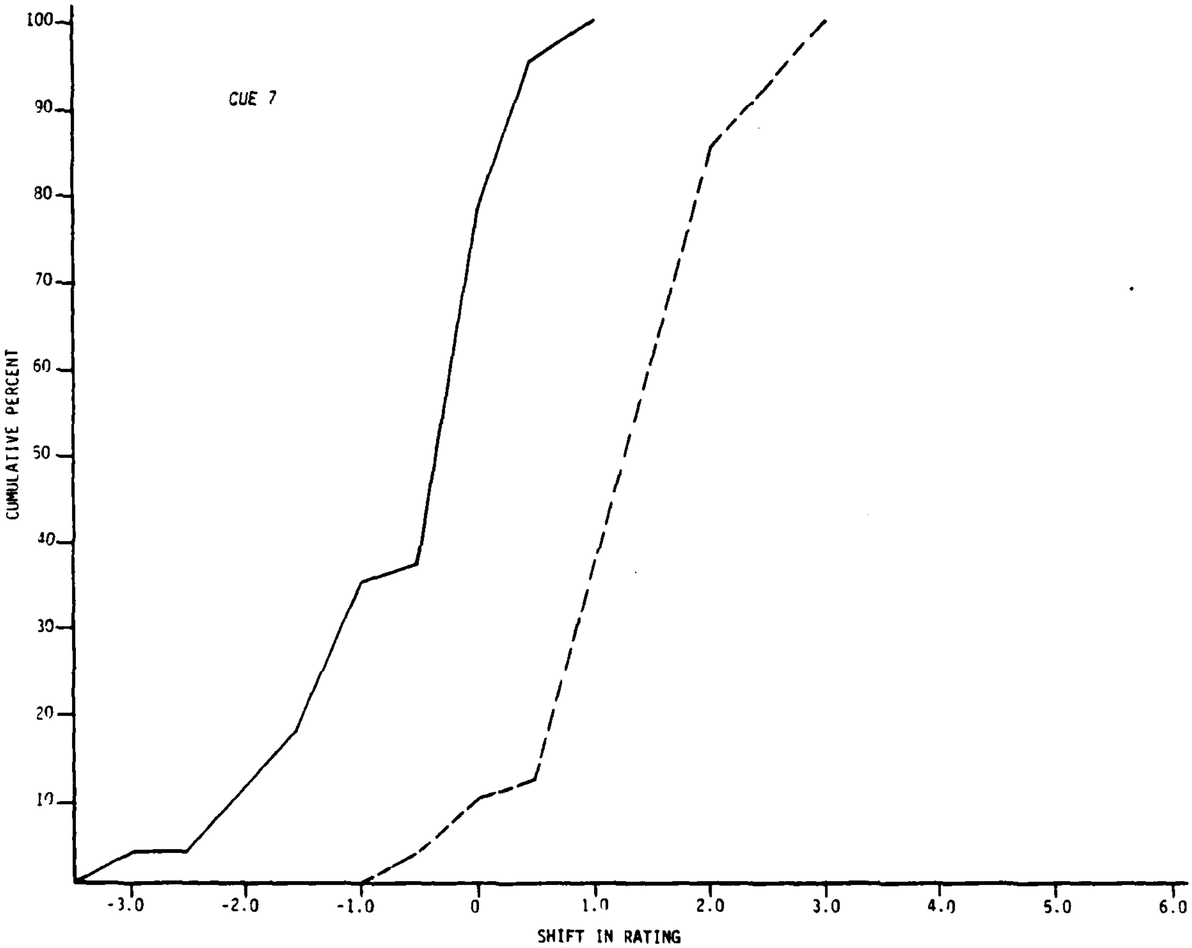


FIGURE 3-14. (CONT.)

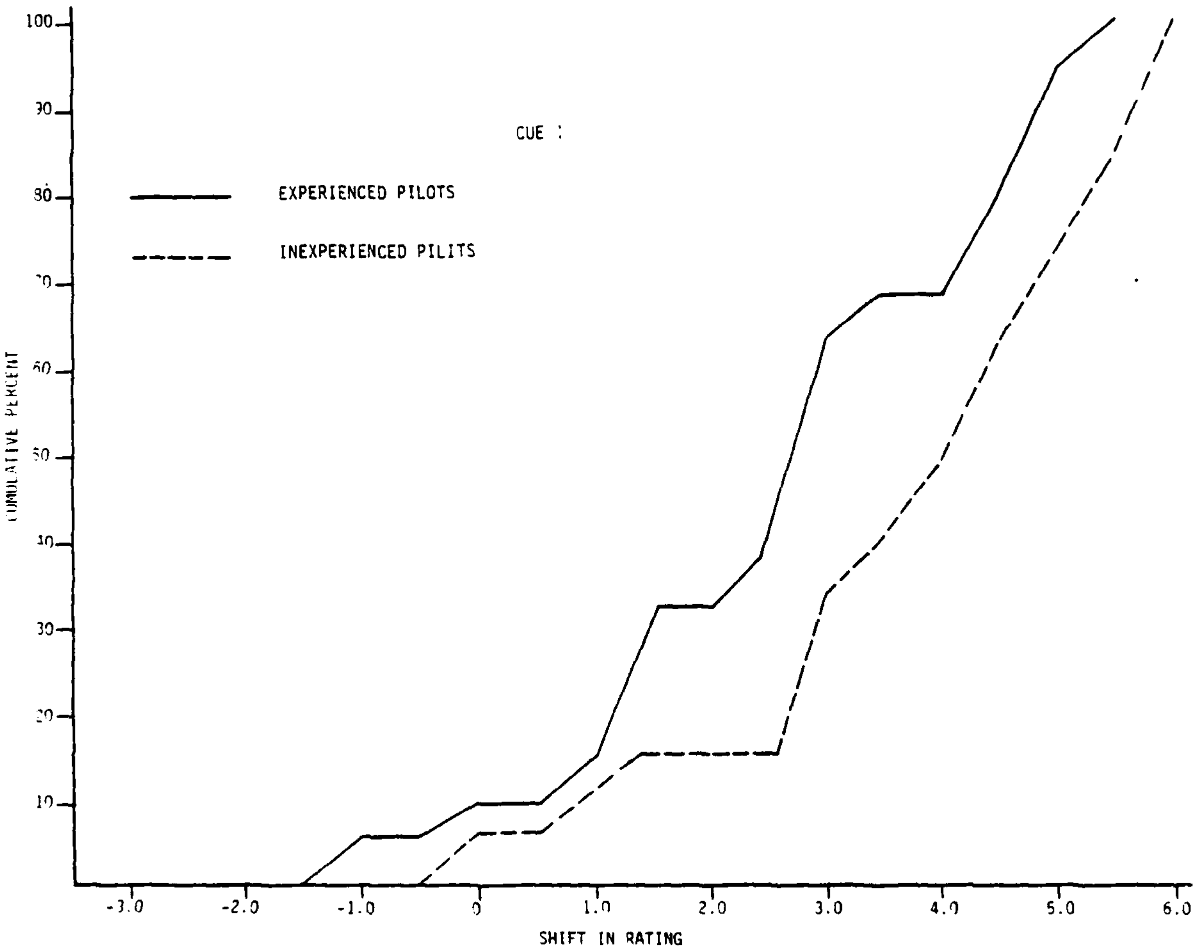


FIGURE 3-15.
SHIFTS IN RATINGS OF SAFETY CRITICALITY BY CUE

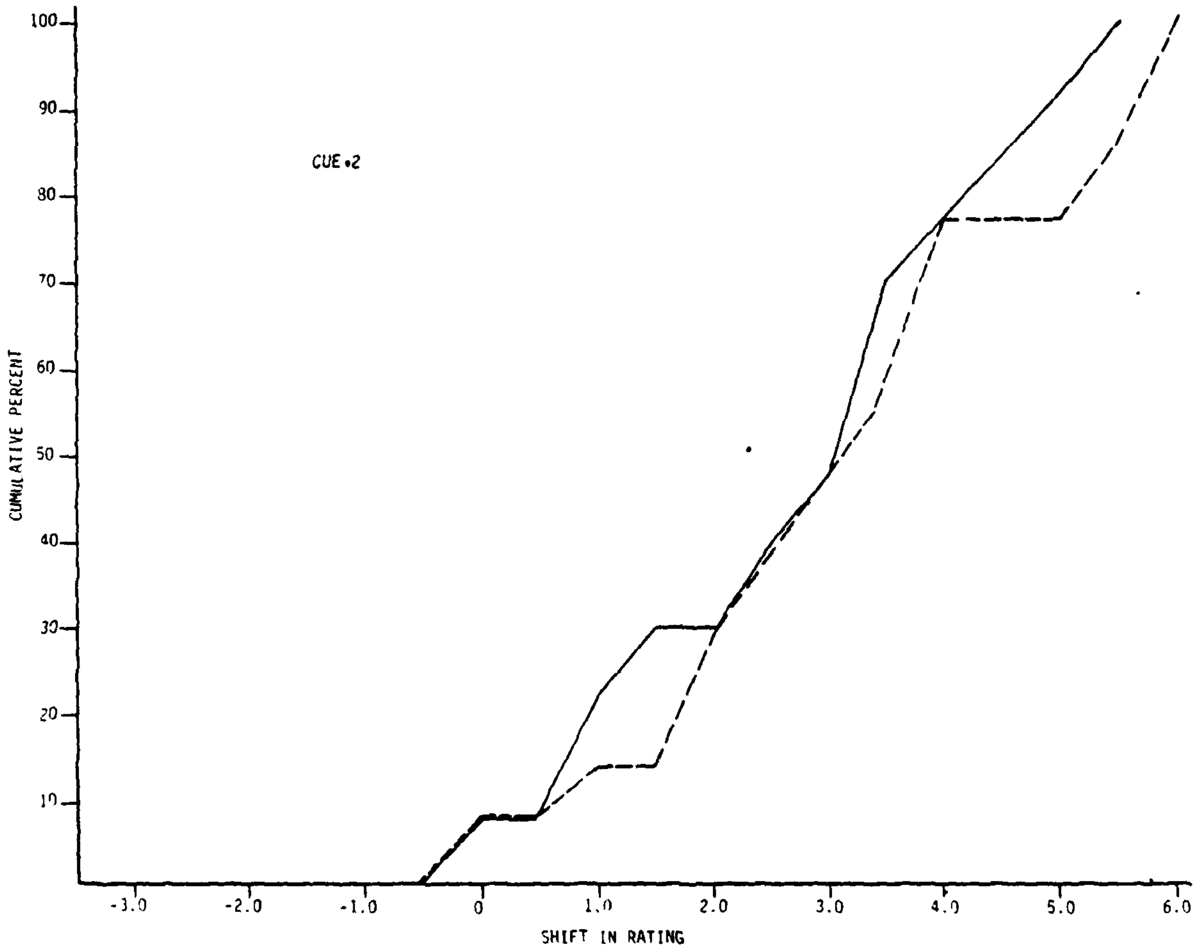


FIGURE 3-15. (CONT.)

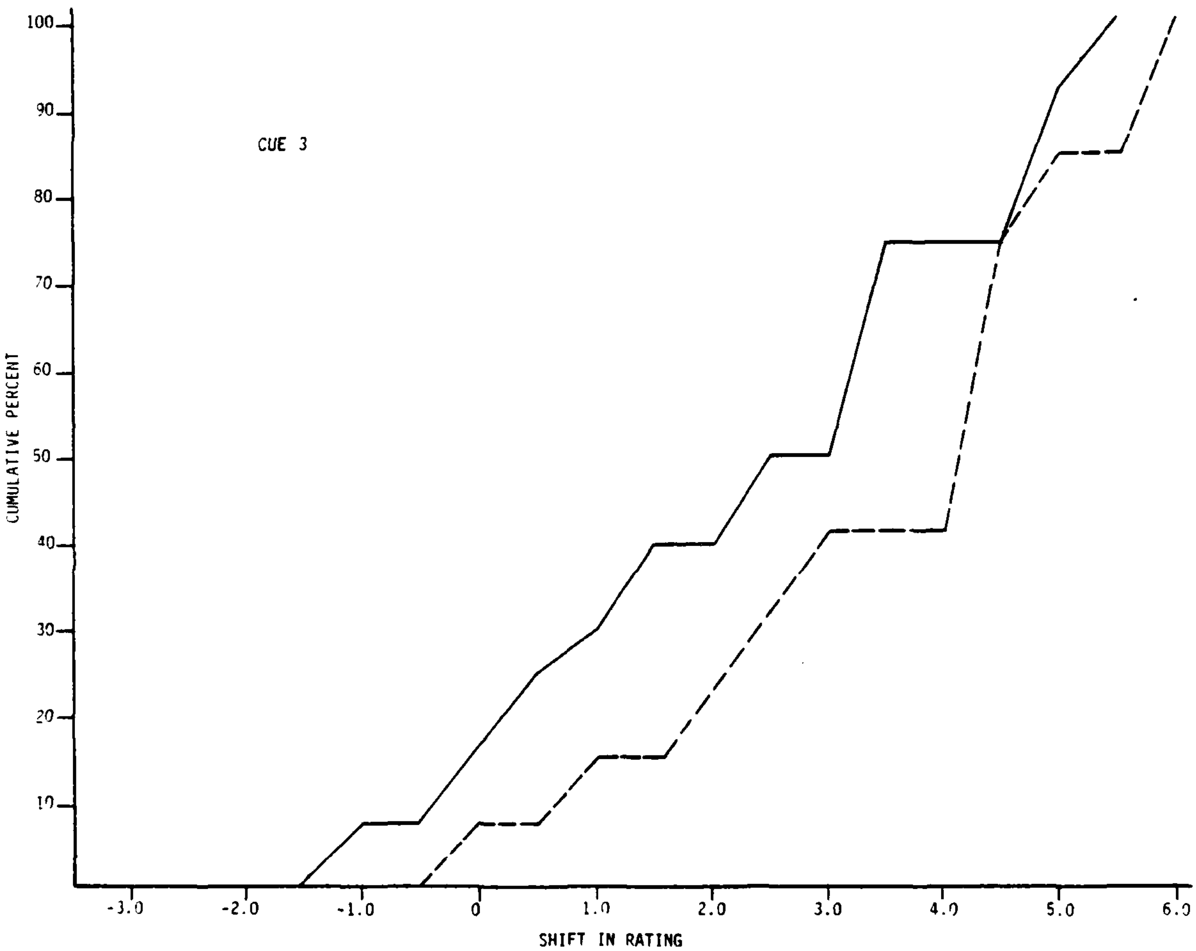


FIGURE 3-15. (CONT.)

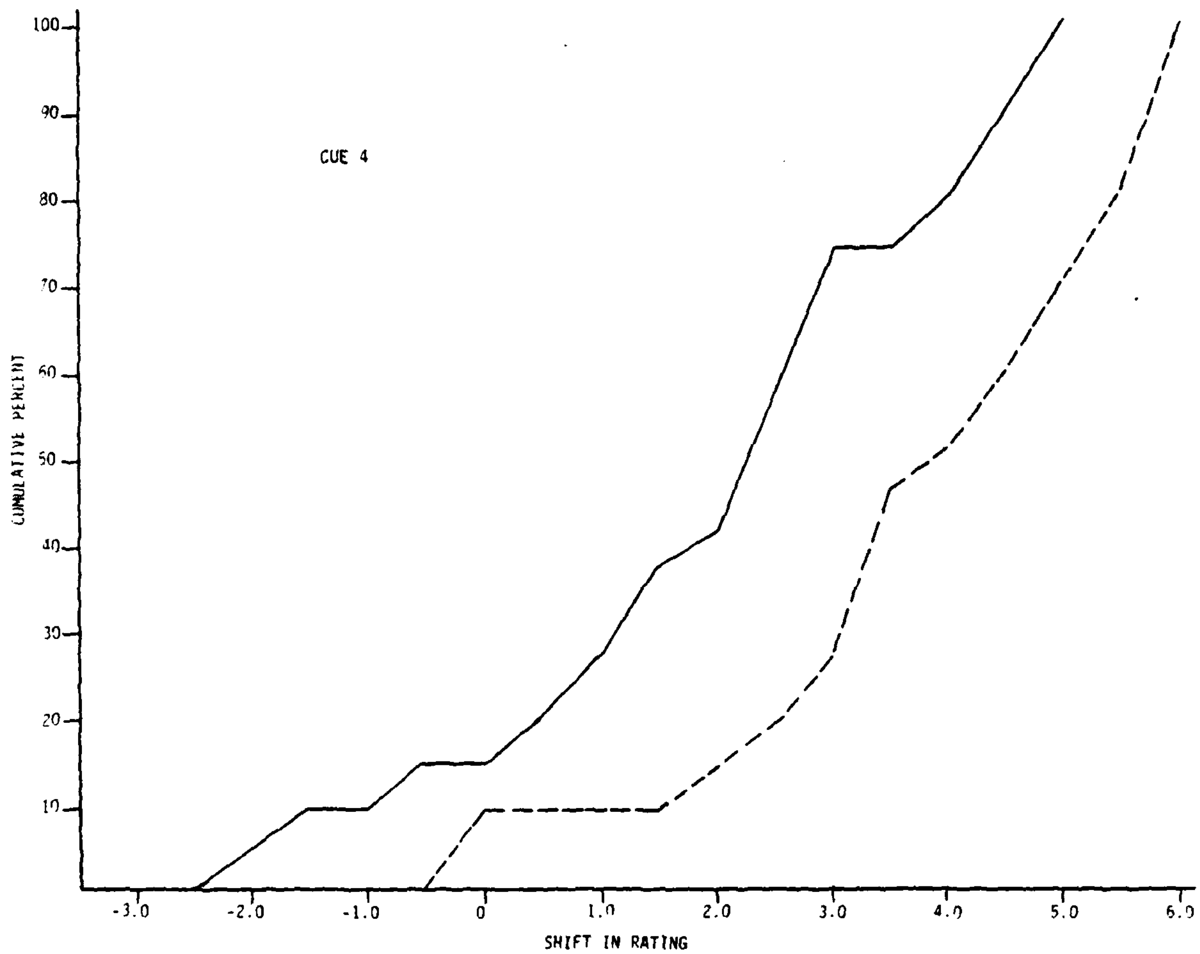


FIGURE 3-15. (CONT.)

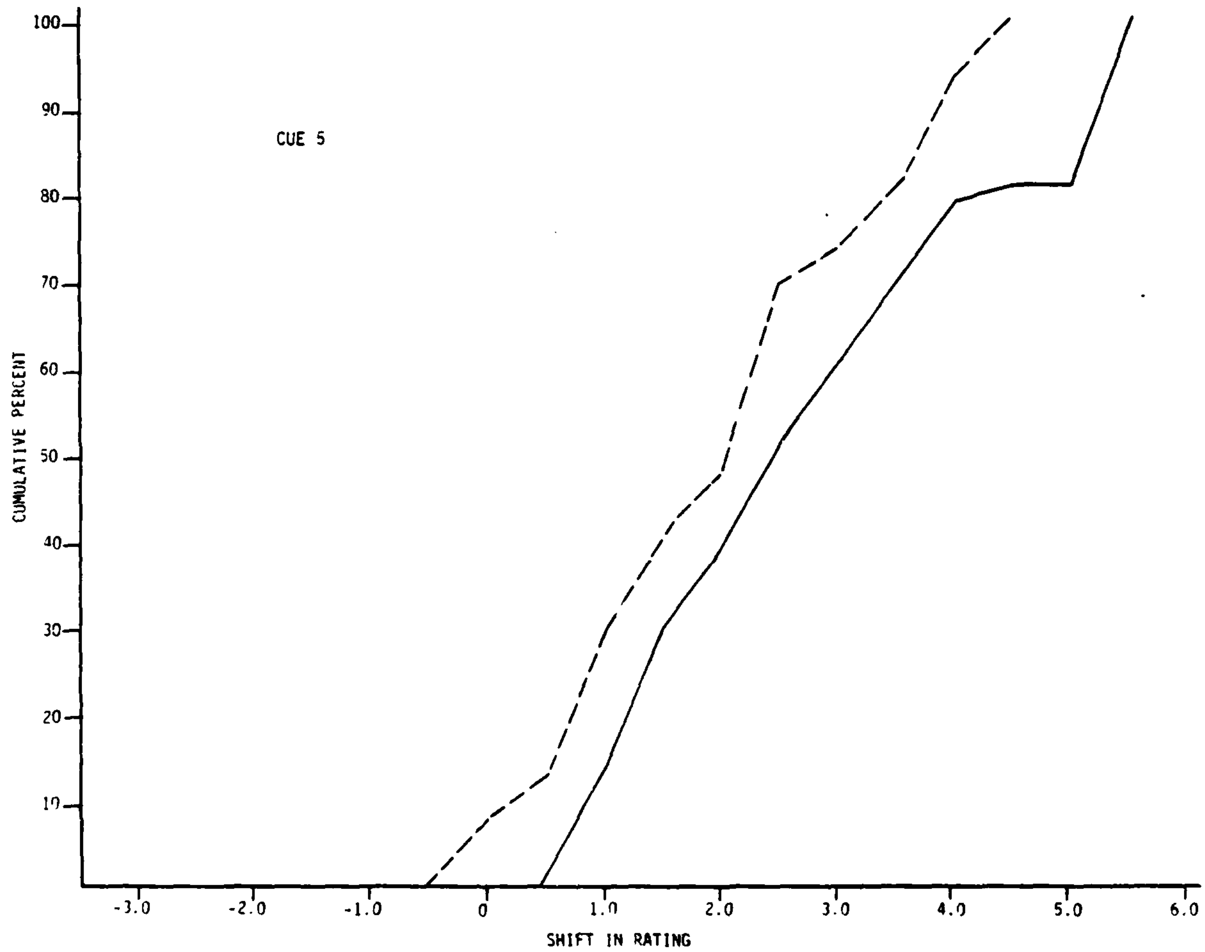


FIGURE 3-15. (CONT.)

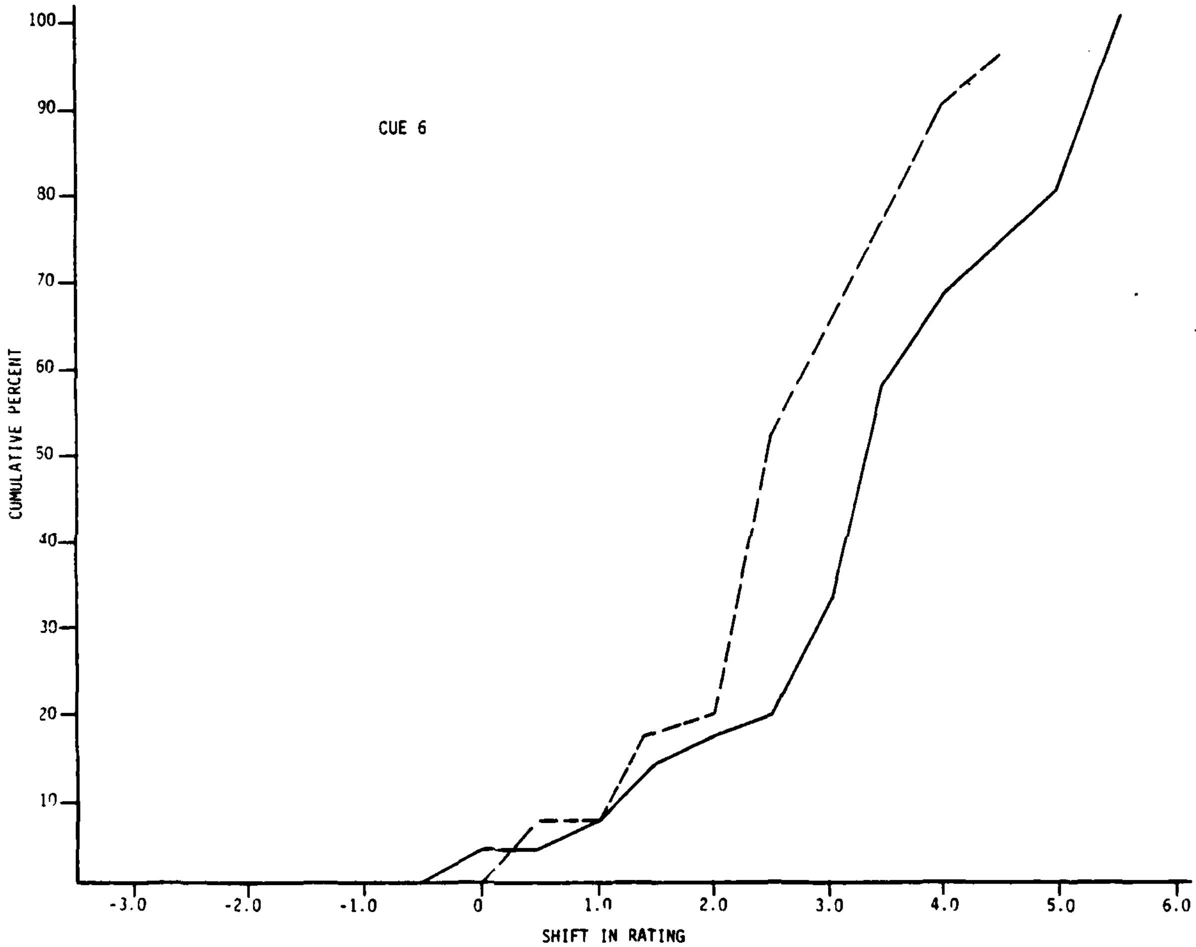


FIGURE 3-15. (CONT.)

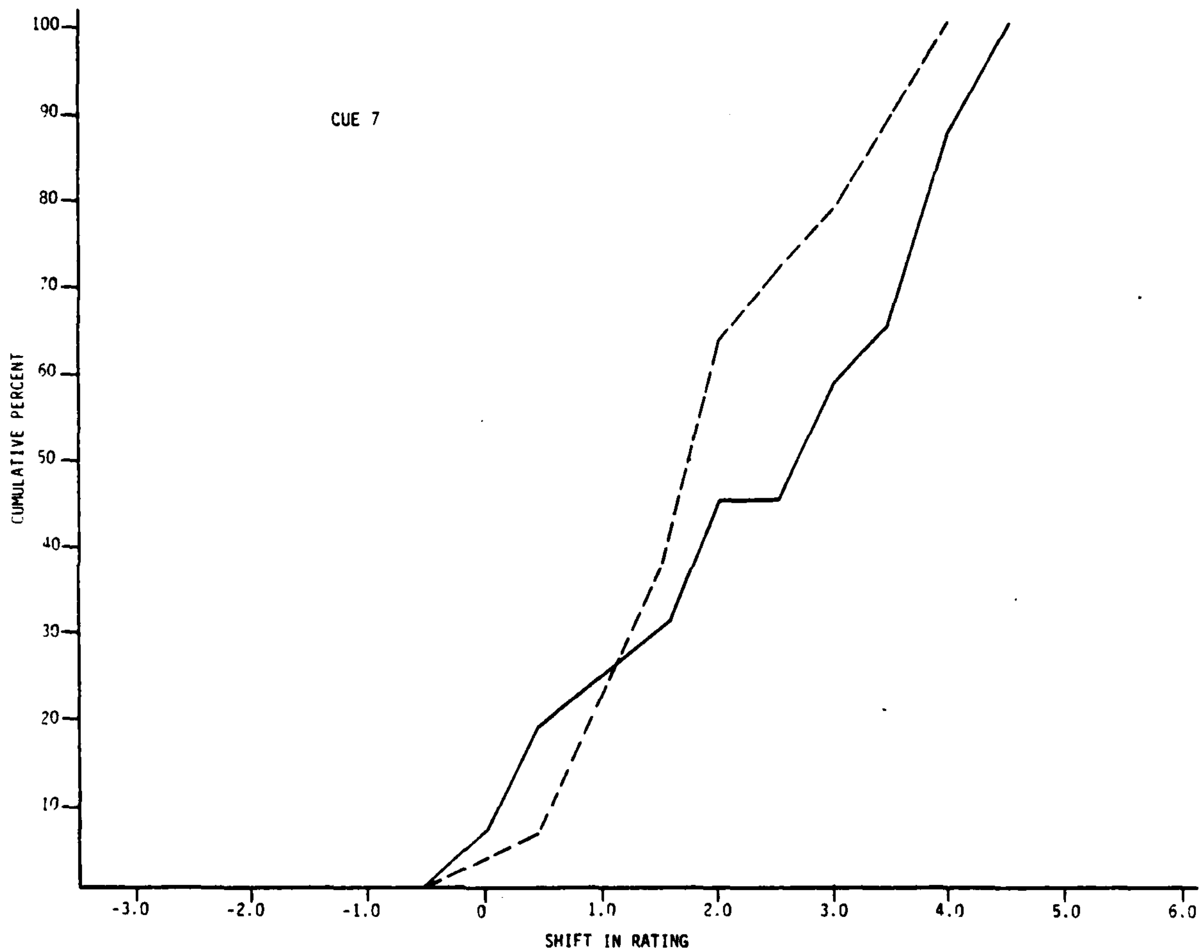


FIGURE 3-15. (CONT.)

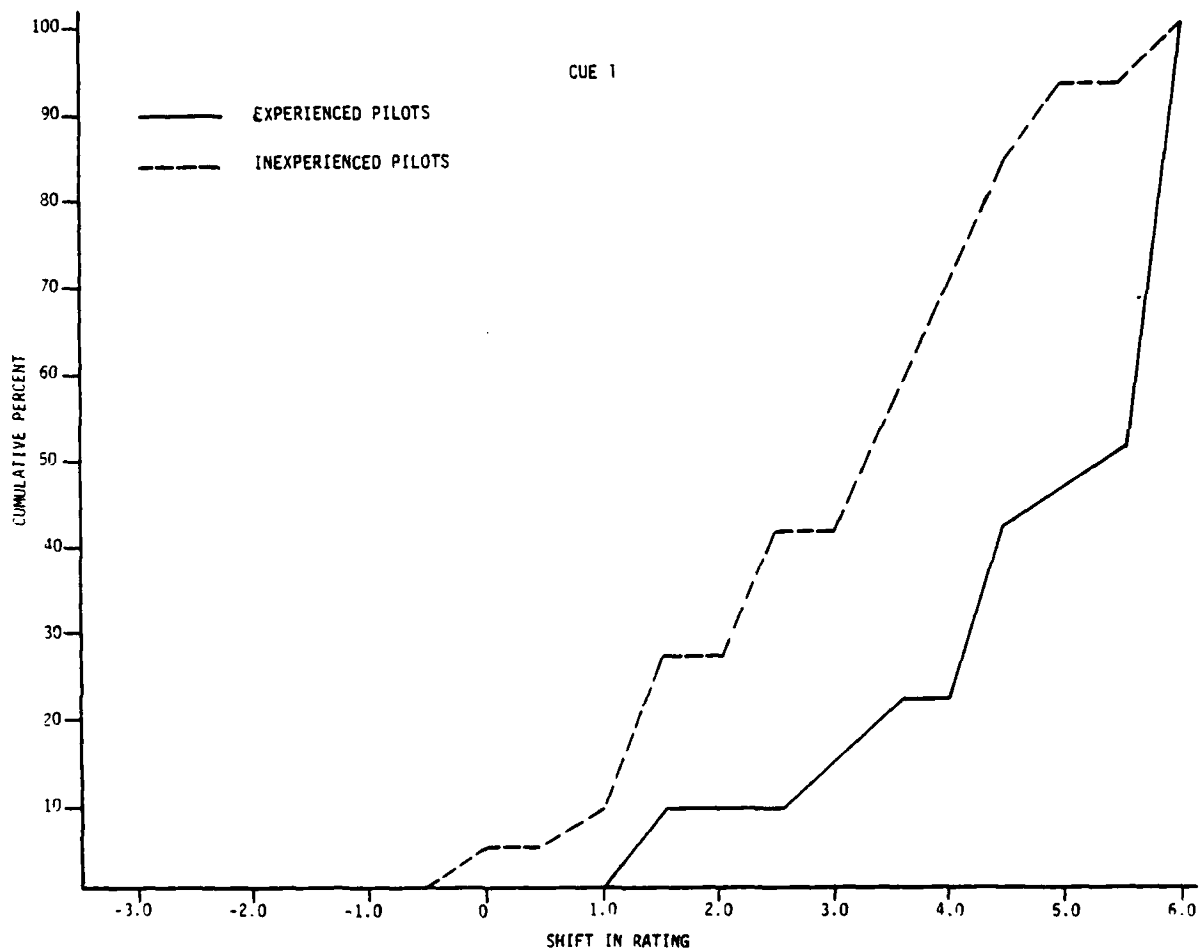


FIGURE 3-16.
SHIFTS IN RATINGS FOR TIME CRITICALITY BY CUE

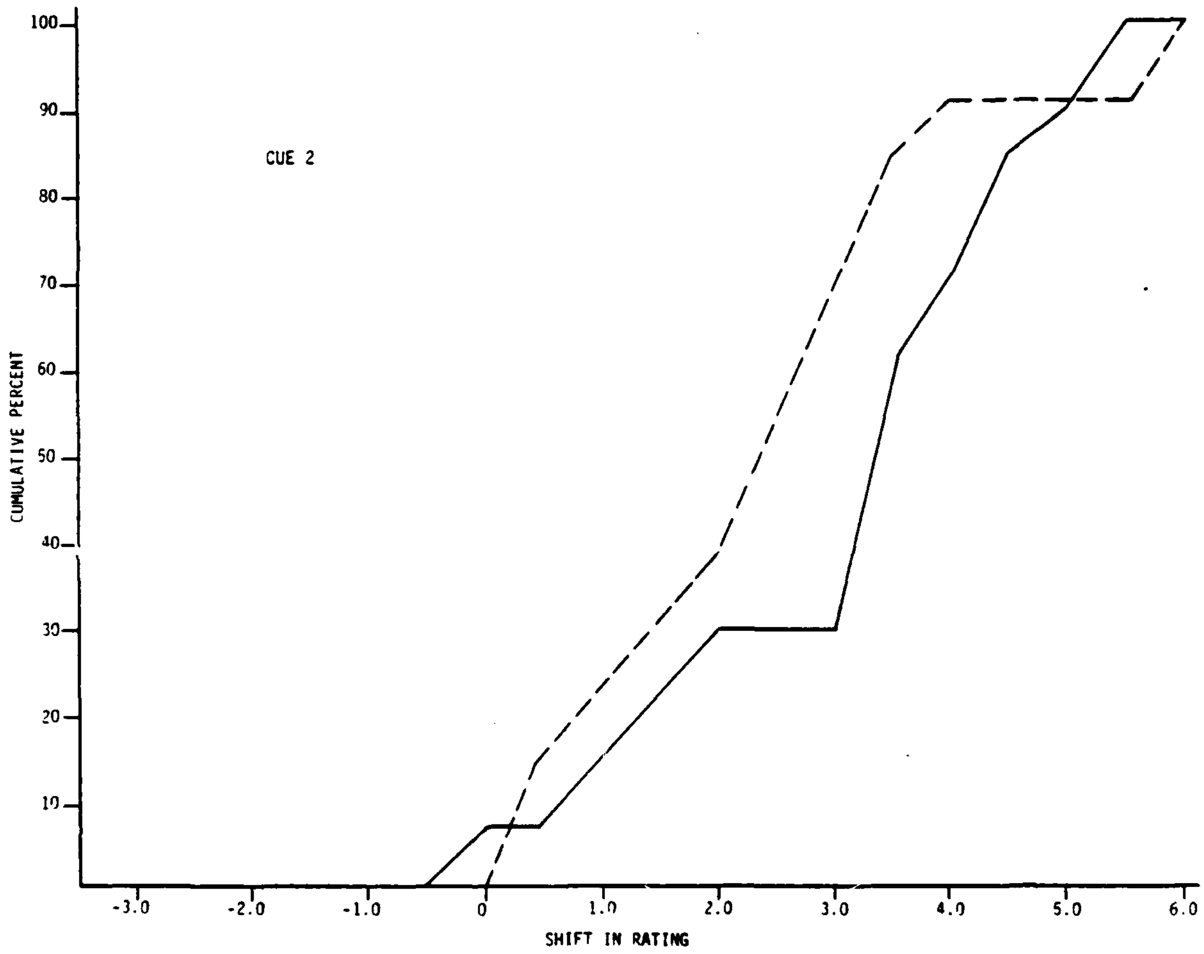


FIGURE 3-16. (CONT.)

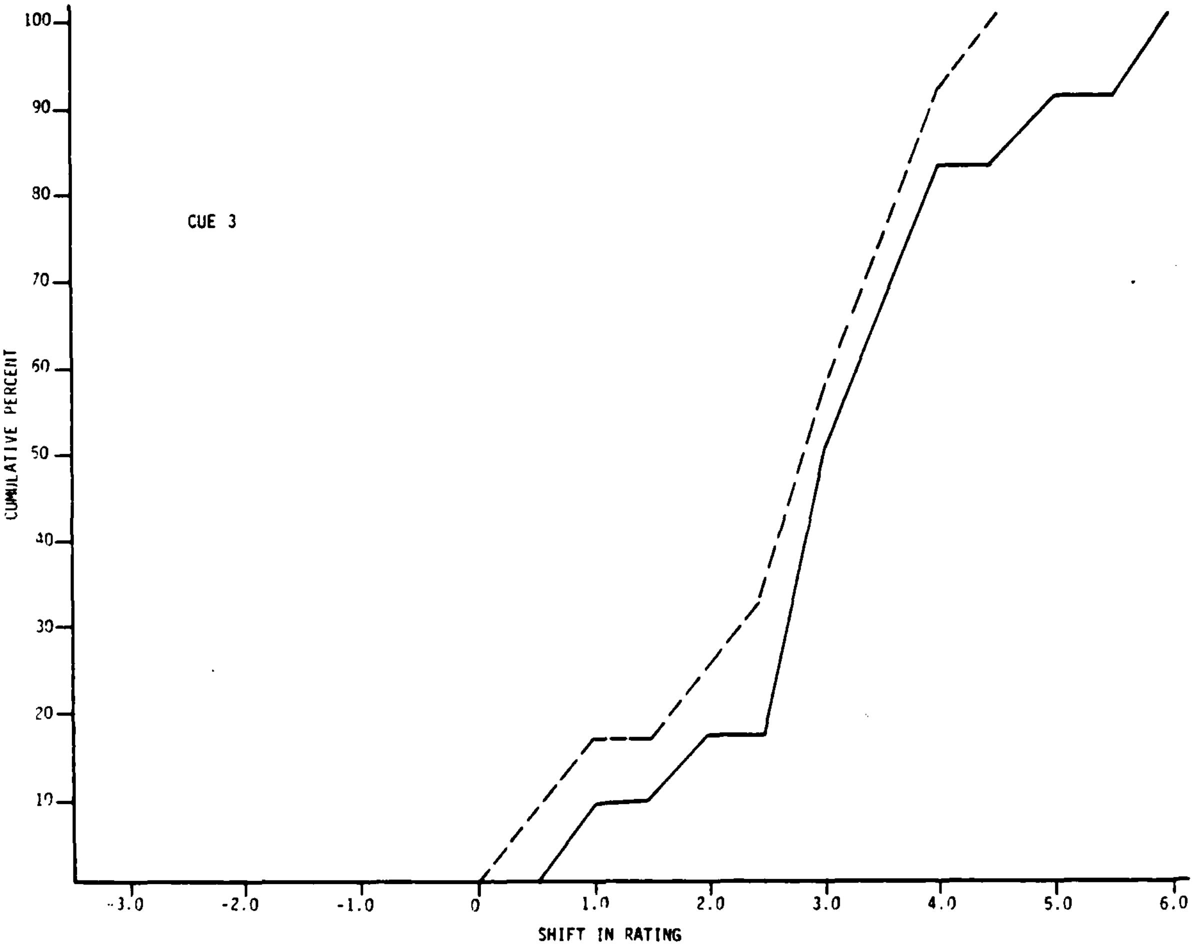


FIGURE 3-16. (CONT.)

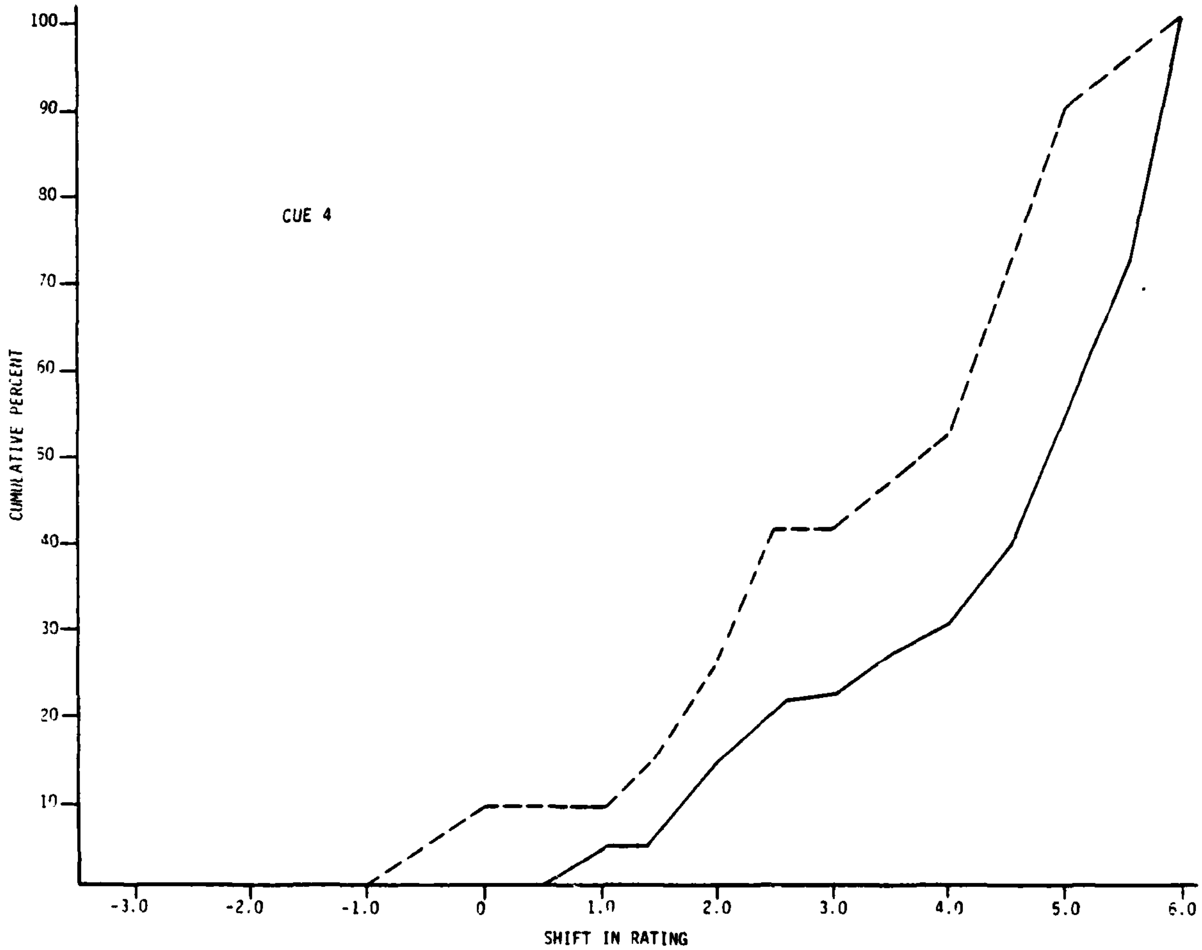


FIGURE 3-16. (CONT.)

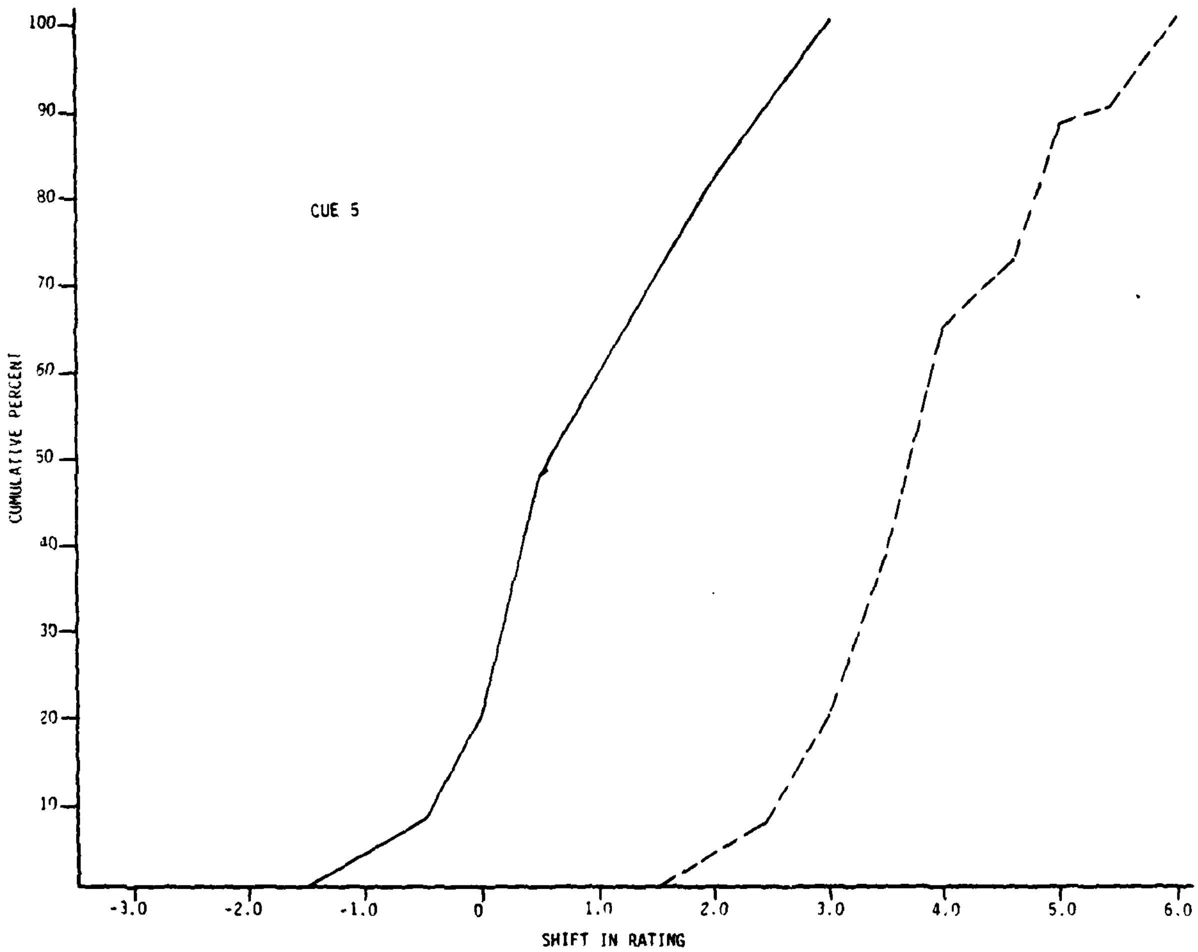


FIGURE 3-16. (CONT.)

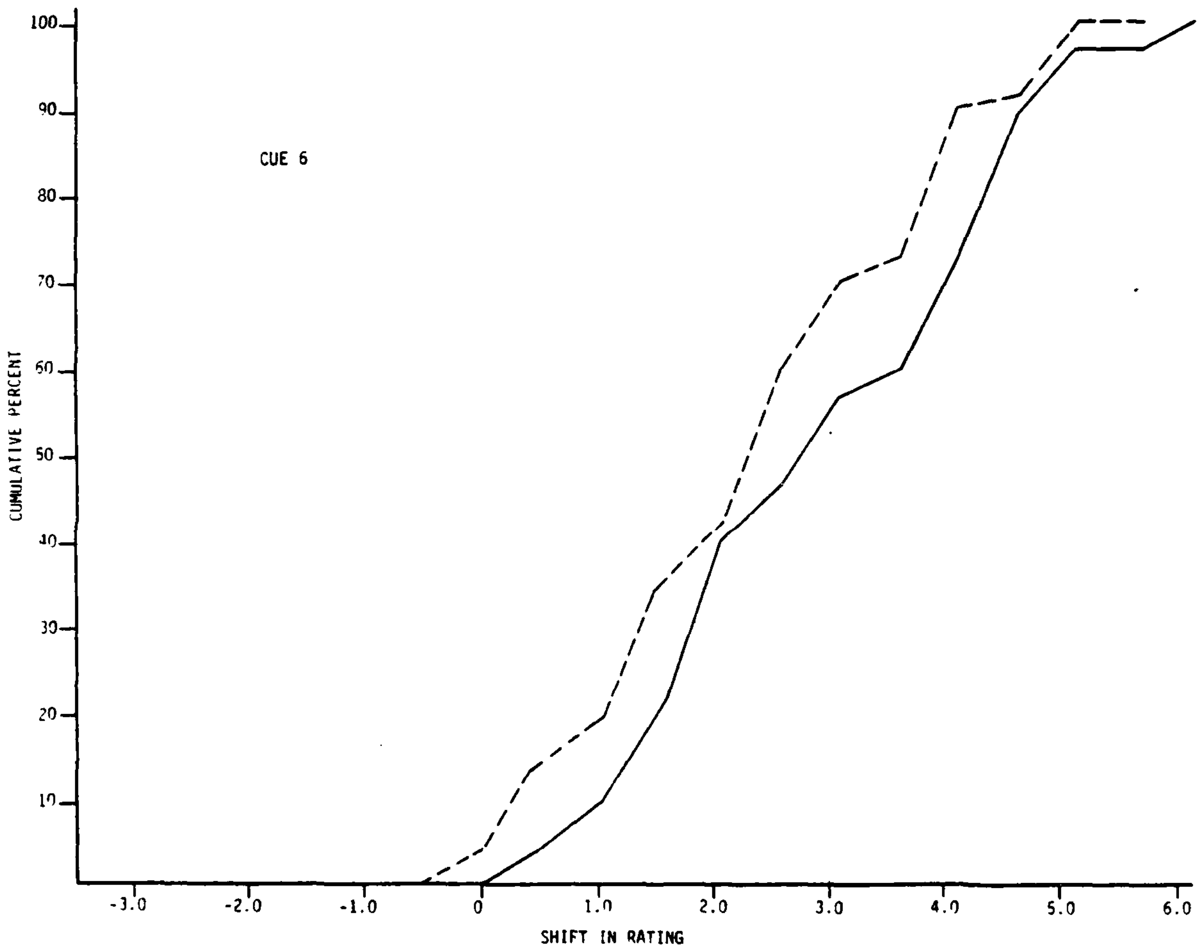


FIGURE 3-16. (CONT.)

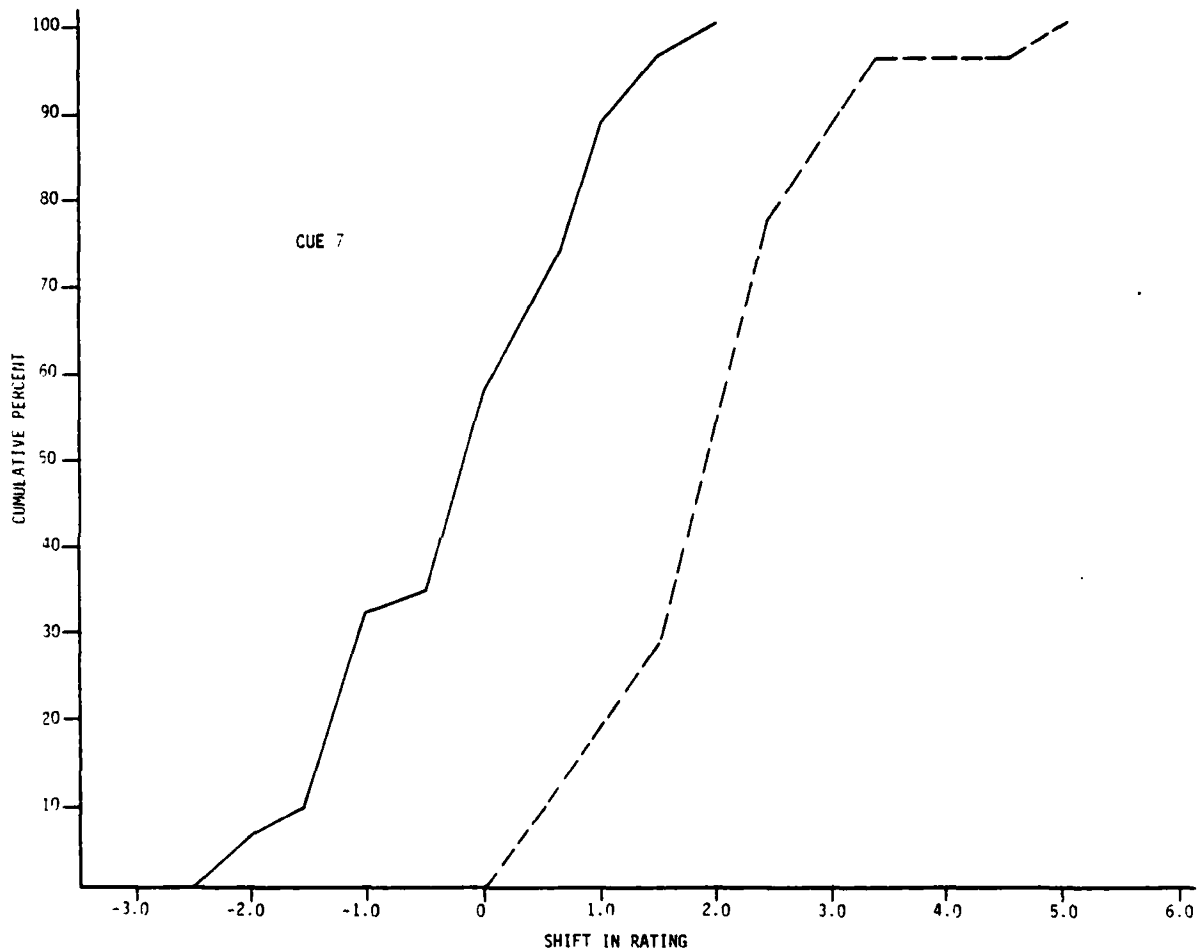


FIGURE 3-16. (CONT.)

Figure 3-17 summarizes the impact of the addition of each of the two situation cues (weather, flight phase) on attribute judgments. The impact of each type of cue was found to differ depending on the number of other (malfunction-related) cues present. The factor of weather seems to produce significant shifts in judgment in about 42% of the cases when there is only one additional malfunction-related cue; as the number of malfunction-related cues was increased, however, weather tends to be ignored as evidenced by fewer shifts in cue pattern rating. This is not the case for flight phase, where the shift in ratings remains relatively constant regardless of the number of other cues present. This difference suggests further that information load per se is not a controlling factor, but that the nature of the situational cue (weather vs. flight phase) is critical. The data in Figure 3-17 are pooled for both experience levels since both groups of pilots exhibited the same results. (As an aside, the apparent tendency shown here to ignore weather as a variable in judgment situations involving several other cues, may have interesting implications for the relatively high number of aircraft accident reports which identify weather as a contributing factor.)

The data in Figure 3-17 are averaged over all attributes, cues and experience levels. More detailed analyses, not shown here, do indicate that situational cues can be treated similarly to malfunction cues in terms of discriminability and consistency of ratings and in terms of impact, given the differences noted above.

3.6 Summary

The results of these studies indicate that both malfunction-related and situational cues can be assessed by flying personnel in a consistent fashion and that the ratings of both such cues appear to result from a similar process. Ratings were obtained from both experienced and inexperienced pilots. While the ratings appeared to differ widely for some

	NUMBER OF MALFUNCTION CUES		
	1	2	3 or 4
WEATHER	42%	14%	20%
FLIGHT PHASE	43%	36%	42%

FIGURE 3-17.
PERCENT RESPONSES INDICATING EFFECT OF SITUATIONAL CUES ON RATINGS AS A FUNCTION OF THE NUMBER OF MALFUNCTION CUES IN PATTERNS

individual cues, striking differences did not consistently emerge between the two groups for all combinations of cues and attributes. A particularly interesting feature of the data obtained was the identification of cues which appear to remain dominant when configured with other cues into cue patterns, and the identification of cues whose ratings do not remain stable when embedded in patterns. This methodology may provide a means to identify information processing problems and/or learning difficulties associated with particular situations (cue combinations) that can be used for the design of synthetic learning environments for emergency decision training.

4. CONSIDERATIONS FOR EVALUATING DECISION TRAINING

4.1 Introduction

As an adjunct to the research studies of Year 2, a review of relevant evaluation theory and findings was conducted to aid those who might consider testing improved scenario design procedures for aircrew emergency training. This Chapter summarizes the results of the literature review according to the following categories: (a) meta-evaluation theory, (b) evaluation principles and concerns, (c) technical discussions and research findings related to the evaluation of decision training, and (d) an overview of evaluation considerations particularly relevant to evaluating aircrew emergency decision skills and skill training.

4.2 Meta-Evaluation

The term "meta-evaluation" was coined by Scriven (1969) to refer to the evaluation of evaluation systems, specifically to assessing the adequacy of proposed designs for evaluation projects and completed studies. Eleven criteria are proposed by Stufflebeam (1974) for judging the merits of evaluation designs and studies. They are grouped below under these general headings--technical adequacy, usefulness, and cost-effectiveness:

Technical Adequacy

(1) Internal validity--Are the results obtained accurate and true? Do they answer the questions asked?

(2) External validity--Can the results obtained be generalized to external settings of interest?

(3) Reliability--Are the results obtained stable and repeatable?

(4) Objectivity--Have sources of bias been controlled?

Utility

(5) Relevance--Is the evaluation designed to answer questions of interest to the decision makers who are the primary audience(s) of the study?

(6) Importance--Does the evaluation focus on collecting those data which best serve decision makers?

(7) Scope--Does the evaluation address all important areas of interest?

(8) Credibility--Do the conditions of evaluation generate confidence in the results obtained?

(9) Timeliness--Are results obtained early enough to influence decisions of interest, particularly formative decisions?

(10) Pervasiveness--Are evaluative results disseminated to all audiences which have a need for them?

Cost-effectivess

(11) Cost-effectiveness--Is the most efficient alternative for conducting the evaluation chosen without sacrificing the quality of the evaluation?

Stufflebeam (1974) points out the difficulty of optimizing a particular evaluation design in light of these factors:

> *In summary, evaluations should be technically adequate, useful, and efficient. The eleven criteria presented above are suggested to meta-evaluators for their use in assessing evaluation designs and reports. It is apparent that the evaluator cannot insist on optimizing any one criterion if he is to optimize his over-all effort. Rather he must make many compromises and strike the best balance he can in satisfying standards of technical adequacy, utility, and cost/effectiveness. (p.11)*

Figure 4-1 illustrates how these criteria can be integrated into a matrix of questions to review goals, design elements, implementation plans, and results of evaluation. Figure 4-2 illustrates a framework for assessing an evaluation system following its specification. Taken together these figures illustrate the body of concerns and levels of relationships that meta-evaluation deals with.

Since meta-evaluation is a form of evaluation itself, its conceptualization is founded on the eight evaluation premises listed below (Stufflebeam, 1974, pp. 70-71):

(1) Evaluation is the assessment of merit; thus, meta-evaluation means assessing the merit of evaluation efforts.

(2) Evaluation serves decision-making and accountability; thus meta-evaluation should provide information pro-actively to support the decisions that must be made in conducting evaluation work, and meta-evaluation should provide retroactive information to help evaluators be accountable for their past evaluation work. Another way of saying this is that meta-evaluation should be both formative and summative.

(3) Evaluations should assess goals, designs, implementation, and results. Thus meta-evaluation should assess the importance of evaluation objectives, the appropriateness of evaluation designs, the adequacy of implementation of the designs, and the quality and importance of evaluation results.

(4) Evaluation should provide descriptive and judgmental information and appropriate recommendations. Likewise, meta-evaluation should describe and judge evaluation work and should recommend how the evaluations can be improved and how the findings can appropriately be used.

Purposes of Evaluation Studies	CATEGORIES OF EVALUATIVE QUESTIONS			
	Goals	Designs	Implementation	Results
Pro-active Evaluation to serve Decision Making	Who is to be served? What are their needs? What problems have to be solved if the needs are to be met? What funds are available for work in this area? What research findings have a bearing on problem solving in this area? What relevant technology is available? What alternative goals might be chosen?	Are the given objectives stated operationally? Is their accomplishment feasible? What relevant strategies exist? What alternative strategies can be developed? What are the potential costs and benefits of the competing strategies? What are the operating characteristics of the competing strategies? How compatible are the competing strategies with the system? How feasible are the competing strategies?	What is the schedule of activities? What are the personnel assignments? What's the program budget? What potential problems attend the design? What are the discrepancies between the design and the operations? What design changes are needed? What changes in implementation are needed?	What results are being achieved? Are they congruent with the objectives? Are there any negative side effects? Are there any positive side effects? Do the results suggest that the goals, designs, or process should be modified? Do the results suggest that the project will be a success?
Retroactive Evaluation to serve Accountability	What goals were chosen? What goals were considered, then rejected? What alternative goals might have been considered? What evidence exists to justify the goals that were chosen? How defensible is this evidence? How well have the goals been translated into objectives? Overall, what is the merit of the goals that were chosen?	What strategy was chosen? What alternative strategies were considered? What other strategies might have been considered? What evidence exists to justify the strategy that was chosen? How defensible is this evidence? How well was the chosen strategy translated into an operational design? Overall, what is the merit of the chosen strategy?	What was the operational design? To what extent was it implemented? What were the strengths and weaknesses of the design under operating conditions? What was the quality of the effort to implement it? What was the actual design that was implemented? Overall, what is the merit of the process that was actually carried out?	What results were achieved? Were the stated objectives acheived? What were the positive and negative side effects? What impact was made on the target audience? What long term effects may be predicted? What is the relation of costs to benefits? Overall, how valuable were the results and impacts of this effort?

FIGURE 4-1.
A MATRIX FOR IDENTIFYING AND ANALYZING EVALUATIVE QUESTIONS

Purposes of the Meta-Evaluation	Steps in the Meta-Evaluation Process	Objects of the Meta-Evaluation			
		Evaluation Goals	Evaluation Designs	Evaluation Process	Evaluation Results
Pro-active Meta-Evaluation to serve Decision Making in eval. work (This is formative Meta-Evaluation and usually is conducted by insiders)	Delineating the information requirements	Audiences Possible eval. goals Criteria for rating eval. goals	Alternative eval. designs Criteria for rating eval. designs	Work breakdown and schedule for the chosen eval. design Admin. checklist for reviewing eval. designs	The eval. objectives Cost, quality, and impact criteria Intended users of the evaluation
	Obtaining the needed information	Logical analyses of the eval. goals Ratings of the eval. goals	Ratings of the alternative designs	Review of the eval. design Monitoring of the eval. process	Ratings of the quality of reports Evidence of use of eval. for decision making & accountability Ratings of the value of eval. reports Monitoring of expenditures for eval.
	Applying the obtained information	Recommendations of what eval. goals should be chosen	Recommendations of what eval. design should be chosen	Periodic progress & exception reports Recommendations for modifying the eval. design or procedures	Periodic reports of the quality, impact, & cost/effectiveness of the eval. Recommendations for improving eval. results
Retroactive Meta-Evaluation to serve Accountability in eval. work (This is Summative Meta-Evaluation and usually is conducted by outsiders)	Delineating the information requirements	Audiences Goals chosen Criteria for judging eval. goals	The chosen design The critical competitors Criteria for rating eval. designs	Work breakdown & schedule for the chosen eval. design Admin. checklist for reviewing eval. designs	The eval. objectives Cost, quality, & impact criteria Intended users of the evaluation
	Obtaining the needed information	Survey of evaluation needs Audience ratings of chosen eval. goals Analysis of eval. goals related to criteria, needs, & audience ratings	Ratings of the alternative eval. designs	Case study of the eval. process Analysis of discrepancies between the eval. process & the chosen design	Ratings of the quality of reports Evidence of use of eval. for decision making & accountability Ratings of the value of eval. reports Cost analysis for the evaluation
	Applying the obtained information	Judgment of the chosen eval. goals	Judgment of the chosen eval. design	Judgment of the implementation of the eval. design	Judgment of the quality, utility, and cost/effectiveness of the eval. activity

FIGURE 4-2.
A META-EVALUATION FRAMEWORK

(5) Evaluation should serve all persons who are involved in and affected by the program being evaluated; hence, meta-evaluation should serve evaluators and all the persons who are interested in their work.

(6) Evaluation should be conducted by both insiders and outsiders; generally (but not always) insiders should conduct formative evaluation for decision making, and outsiders should conduct summative evaluation for accountability. Hence, evaluators should conduct formative meta-evaluation and they should obtain external judgments of the overall merit of their completed evaluation activities.

(7) Evaluation involves the process of delineating the questions to be addressed, obtaining the needed information, and using the information in decision making and accountability. Hence, meta-evaluators must implement three steps. The meta-evaluators must delineate the specific meta-evaluation questions to be addressed. They must next collect, organize, and analyze the needed information. Ultimately they must apply the obtained information to the appropriate decision-making and accountability tasks.

(8) Evaluation must be technically adequate, useful, and cost/effective, and meta-evaluation must satisfy the same criteria.

Further, according to the Stufflebeam model, deficiencies or problems of evaluation systems fall into several categories:

(1) Conceptual.
(2) Sociopolitical.
(3) Contractual/Legal.

(4) Technical Design.
(5) Administrative/Management.
(6) Moral/Ethical.

As an outgrowth of the meta-evaluation framework chosen and the premises on which meta-evaluation is based, proposed or actual evaluation systems can be assessed according to the six major categories of concern identified above. Table 4-1 illustrates an administrative checklist developed under the meta-evaluation framework to assess an educational evaluation system.

While meta-evaluation remains a relatively primitive art, Stufflebeam has provided a conceptual structure which can be used in evaluation design to ensure that major design questions are at least considered, even though prescriptive solutions to many design problems may not yet be available.

4.3 Evaluation Principles and Concerns

A number of comprehensive reviews of evaluative techniques and principles are available (e.g., Campbell and Stanley, 1963; Kerlinger, 1964; Moursund, 1973; Popham, 1974; Tien, 1979; and Weiss, 1972). The bulk of the literature available deals with evaluation as research. Although all research is not evaluation, all good evaluation must meet the same requirements that good research must meet.

Reynolds, Neuse, Olds and Levine (1978) and Tien (1979) present particularly comprehensive overviews of the factors to consider in establishing an evaluation design for a particular product. Reynolds *et al.* (1978) provide a process model for systematic consideration of relevant topics necessary to specification of an evaluation design. Chapanis (1958) and Parsons (1972) provide good treatments of the human factors or man-machine aspects of such evaluations.

TABLE 4-1

AN ADMINISTRATIVE CHECKLIST FOR REVIEWING EVALUATION PLANS

Conceptualization of Evaluation

_____Definition	--How is evaluation defined in this effort?
_____Purpose	--What purpose(s) will it serve?
_____Questions	--What questions will it address?
_____Information	--What information is required?
_____Audiences	--Whom will be served?
_____Agents	--Who will do it?
_____Process	--How will they do it?
_____Standards	--By what standards will their work be judged?

Sociopolitical Factors

_____Involvement	--Whose sanction and support is required, and how will it be secured?
_____Internal communication	--How will communication be maintained between the evaluators, the sponsors, and the system personnel?
_____Internal credibility	--Will the evaluation be fair to persons inside the system?
_____External credibility	--Will the evaluation be free of bias?
_____Security	--What provisions will be made to maintain security of the evaluative data?
_____Protocol	--What communication channels will be used by the evaluators and system personnel?
_____Public relations	--How will the public be kept informed about the intents and results of the evaluation?

Contractual/Legal Arrangements

_____Client/evaluator relationship	--Who is the sponsor, who is the evaluator, and how are they related to the program to be evaluated?
_____Evaluation products	--What evaluation outcomes are to be achieved?

TABLE 4-1 (Cont'd)

AN ADMINISTRATIVE CHECKLIST FOR REVIEWING EVALUATION PLANS

_____Delivery schedule	--What is the schedule of evaluation services and products?
_____Editing	--Who has authority for editing evaluation reports?
_____Access to data	--What existing data may the evaluator use, and what new data may he obtain?
_____Release of reports	--Who will release the reports and what audiences may receive them?
_____Responsibility and authority	--Have the system personnel and evaluators agreed on who is to do what in the evaluation?
_____Finances	--What is the schedule of payments for the evaluation, and who will provide the funds?

<u>The Technical Design</u>

_____Objectives and variables	--What is the program designed to achieve, in what terms should it be evaluated?
_____Investigatory framework	--Under what conditions will the data be gathered, e.g., experimental design, case study, survey, site review, etc.?
_____Instrumentation	--What data-gathering instruments and techniques will be used?
_____Sampling	--What samples will be drawn, how will they be drawn?
_____Data gathering	--How will the data-gathering plan be implemented, who will gather the data?
_____Data storage and retrieval	--What format, procedures, and facilities will be used to store and retrieve the data?
_____Reporting	--What reports and techniques will be used to disseminate the evaluation findings?
_____Technical adequacy	--Will the evaluative data be reliable, valid, and objective?

TABLE 4-1 (Cont'd)

AN ADMINISTRATIVE CHECKLIST FOR REVIEWING EVALUATION PLANS

The Management Plan

_____Organizational mechanism	--What organizational unit will be employed, e.g., an in-house office of evaluation, a self evaluation system, a contract with an external agency, or a consortium-supported evaluation center?
_____Organizational location	--Through what channels can the evaluation influence policy formulation and administrative decision making?
_____Policies and procedures	--What established and/or ad hoc policies and procedures will govern this evaluation?
_____Staff	--How will the evaluation be staffed?
_____Facilities	--What space, equipment, and materials will be available to support the evaluation?
_____Data-gathering schedule	--What instruments will be administered, to what groups, according to what schedule?
_____Reporting schedule	--What reports will be provided, to what audiences, according to what schedule?
_____Training	--What evaluation training will be provided to what groups and who will provide it?
_____Installation of evaluation	--Will this evaluation be used to aid the system to improve and extend its internal evaluation capability?
_____Budget	--What is the internal structure of the budget, how will it be monitored?

Moral/Ethical/Utility Questions

_____Philosophical stance	--Will the evaluator be value free, value based, or value plural?
_____Service orientation	--What social good, if any, will be served by this evaluation, whose values will be served?

TABLE 4-1 (Cont'd)

AN ADMINISTRATIVE CHECKLIST FOR REVIEWING EVALUATION PLANS

_____Evaluator's values	--Will the evaluator's technical standards and his values conflict with the client system's and/or sponsor's values; will the evaluator face any conflict of interest problems; and what will be done about possible conflicts?
_____Judgments	--Will the evaluator judge the program; leave that up to the client; or obtain, analyze, and report the judgments of various reference groups?
_____Objectivity	--How will the evaluator avoid being co-opted and maintain his objectivity?
_____Prospects for utility	--Will the evaluation meet utility criteria of relevance, scope, importance, credibility, timeliness, and pervasiveness?
_____Cost/effectiveness	--Compared to its potential payoff will the evaluation be carried out at a reasonable cost?

(From Stufflebeam, 1974)

Five areas to be reviewed below represent the major concerns in developing an effective evaluation design. They are:

(1) The evaluation topic.
(2) The research design.
(3) Statistical design and data analysis.
(4) Subjects.
(5) Special considerations for testing instructional products.

4.3.1 The Evaluation Topic. Reynolds et al. (1978) point out that as an initial step in evaluation design a clear specification of the topic, problem or issue to be evaluated is necessary. This step should be conducted with the assistance of the ultimate user or decision maker for whom the evaluation results are intended. The purpose is to weed out irrelevant, infeasible or unnecessary topics and to ensure relevance of results to subsequent actions and decisions.

The topic of an evaluation can be an object, person, event, activity, project, or system. One or more topics can be considered. Each can have multiple alternatives which would be compared to each other, or they can be independent topics which are not necessarily comparable. If the subject(s) of the evaluation cannot be clarified through discussion and review of policy and doctrine, an exploratory study may be required to identify major elements and factors for study.

For instructional innovations it is possible to evaluate the innovation only on the basis of its stated performance goals or intended impact. Frequently, however, a comparision to current or traditional practices is required to provide additional perspective on the efficacy of the innovation.

Scriven (1974) feels that there are few points where good evaluators distinguish themselves more clearly than in their choice of critical competitors:

> *Few if any useful evaluations avoid the necessity to present data on the comparative performance of critically competitive products. All too often the data refers to some pre-established standards of merit, and the reader has no idea whether one can do better for less, or twice as well for 5 percent more, which is the kind of information a consumer wants. Where comparisions are done, the results are sometimes useless because the competitor is chosen so as to give a false impression. The worst example of this is the use of a single "no treatment" or "last year's treatment" control group. It is not too thrilling to discover that an injection of $100,000 worth of computer-assisted instruction (CAI) can improve the math performance of a school by 15 percent if there is a possibility that $1,500 worth of programmed texts would do as well or better. (p. 15)*

Once the specific subject(s) of the evaluation, including control or critical comparision subjects, is (are) specified, it is necessary to consider what aspect(s) of the subject should be evaluated. Four general choices exist:

(1) Need.
(2) Design.
(3) Performance.
(4) Impact.

Impact evaluation is, perhaps, the most commonly understood form of evaluation, and is concerned with whether the product or program resulted in the long-range effects desired of it. Performance evaluation refers to the evaluation of whether a product or program meets the specifics of its design. Design evaluation deals with whether a product's or program's design responds to the needs underlying the design. Needs assessment refers to studies of an existing situation or state to identify areas in which modifications or innovations are needed.

It may be apparent from the proceeding that each stage of evaluation can serve as the standard for the succeeding stage. For instance, the evaluation of performance of an advanced fighter aircraft system could be done in terms of its design, and the design adequacy could be assessed against the specific needs which the system is intended to address.

A related aspect of the process of specifying an evaluation subject is to decide whether effectiveness, efficiency or both dimensions of the subject of evaluation are of interest. Effectiveness deals with the nature of the inputs, outputs, effects and impacts of a program or product, while efficiency deals with costs. Frequently both aspects are of interest to decision makers and should be considered.

A final question to consider in specifying the topic of the evaluation, is that of the decisions to be made as a result of the evaluation. These decisions can include: (a) support/no support; (b) intervention/modification; (c) inquiry. There may be multiple decision makers with different information needs (e.g., designers and users). The timing and constraints on decisions are also considerations in determining the feasibility of conducting an evaluation on a specific topic.

In summary, one or more topics can be evaluated, each of which need to be clearly identified. The critical comparison programs or products should be kept in mind if the evaluation is intended to demonstrate the superiority of a new product or program. An evaluation topic can include one or more aspects of the product or program studied: need, design, performance, or impact. Effectiveness and/or efficiency may be of interest. All of the above factors must be considered in light of the needs of those who will be making decisions about the product or program under study and with adequate consideration of the time and other constraints under which the decision makers are operating.

4.3.2 Research Design. Once the topic of evaluation is clearly specified, variables of interest and relationships among them can be identified. Four general categories of variables can be considered, which are treated below in terms of a training system.

Inputs refer to the labor/effort utilized, knowledge and skills of various instructional personnel, materiel required, and costs. Outputs refer to the training services produced such as classes completed, examinations given, and diagnosis and counseling sessions provided. Effects refer to changes in student attitudes, skills, and knowledge. Impacts involve those changes in individual or group capability to meet performance goals in real-world settings.

Hypothesized relationships can exist between all or any of the four categories or stages of the training process. For example, it may be of interest to ask what level of input (time, cost, instructor skill levels) are necessary to improve real-world performance (impact) by 10%. A more sophisticated question may be to ask which combination of training methods results in a given impact at lowest cost.

A number of research designs can be considered for use in answering the questions of interest which are selected from among the many possible relationships among inputs, outputs, effects, and impacts. Campbell and Stanley (1963) and Tien (1979) review non-experimental, quasi-experimental and fully experimental designs. These range in terms of simplicity and expense from one-time surveys and case studies to intensive, long-duration experiments which must meet requirements of precise control of variables and randomization of subject assignment.

The purpose of the evaluation as specified in the selection of the topic of the evaluation and hypothesized relationships among relevant variables will determine whether a rigorous experimental design is required or a less demanding approach will suffice. Reynolds et al. (1978) provide detailed guidance for identifying the research design most appropriate to the purposes of the evaluation.

4.3.3 Statistical Design and Data Collection. The research approach chosen will indicate the nature of the data to be collected and the statistical analyses necessary. Texts on research design and related statistical analyses abound and offer comprehensive guidance (e.g., Edwards, 1960; Kerlinger, 1964; Winer, 1962). A variety of computer-based statistical programs exist which can be utilized to provide descriptive summaries of data, to compute measures of association between variables, and to compute the significance of differences or relationships found.

The most important point to bear in mind in selecting data to be collected is that it should be targeted to answer just those questions identified in the research design. The data collection plan should be checked against all variables and relationships of interest, ensuring that all variables have reliable, valid and practical means of measurement and that appropriate analyses to describe relationships are planned. In this way, the common errors of collecting unnecessary or marginally useful data, of failing to collect all relevant data, and of collecting data in a form that is difficult or impossible to analyze as intended can be avoided.

4.3.4 Subjects. The selection of human subjects to serve in evaluation studies raises a number of questions. If the purpose of the evaluation is for initial design or development of training products, a limited sample may be adequate and most economical. Pipe (1965) describes single-student and small group procedures for testing educational products for purposes of making formative revisions. The need for representativeness to the target population of such subjects is obvious.

For more experimental approaches involving statistical tests of significance, Cohen (1969) provides a methodology for selecting an efficient sample size. His method considers: (1) the statistical test to be used, (2) the significance level to be applied, (3) the minimally acceptable power level, and (4) the anticipated effect size. By application of this procedure, it is possible to determine the minimum sample size necessary to detect a given effect size with reasonable power. At the same time unnecessarily large (and expensive) samples are avoided, as is the finding of significant, but trivial differences between treatments which are attributable to excessive sample sizes.

Chapanis (1958) describes some of the considerations associated with selecting subjects. One important aspect is randomization of subjects, which may be a requirement of the statistical analyses proposed. In educational settings this requirement may be difficult to meet if intact groups (classes) are available only. In some cases, it may be necessary to randomize the assignment of intact groups to treatments rather than the assignment of individual subjects. If this is the case, revisions to the design and statistical tests will be required.

4.3.5 Additional Concerns in Evaluating Prototype Instructional Systems. In addition to the design considerations discussed immediately above, four concerns seem to be of particular importance to projects in which prototype instructional systems are developed for field test and potential field implementation. These four concerns are:

(1) Formative and summative uses of evaluative data.
(2) Balancing laboratory and field testing.
(3) Maintaining objectivity and credibility in evaluations where the developer is the evaluator.
(4) Ensuring that both instructor and student user needs are considered.

Each of these topics is treated separately below.

4.3.5.1 Formative and Summative Evaluation. Scriven (1967) distinguishes between evaluation designed to improve or modify a product in its developmental stages (formative evaluation) and evaluation designed to appraise a product which has been completed and is in use or ready for use (summative evaluation). The distinction deals with the purpose or goal of the evaluation rather than with the techniques used.

Formative evaluation used to improve an educational product or innovation while it is still in a fluid, developmental state, may do more to ensure a useful end-product than carefully designed, extensive summative evaluation studies of impact and merit, carried out after design and development work is completed (Scriven, 1967). Any responsible product developer will therefore conduct a mix of formative and summative assessments, balanced so as to provide adequate guidance during product development and to obtain convincing evidence of the merit of the completed product during final testing. From a methodological viewpoint, formative and summative evaluation are similar. Both must meet meta-evaluation critieria of technical adequacy (i.e., internal validity, external validity, reliability, and objectivity). Both, therefore, depend on proper application of research design principles and on appropriate methodological procedures for collection, analysis and interpretation of data. In some cases, formative and summative evaluation data are identical, distinguishable only in terms of the uses to which they are put (i.e., the evaluation questions they are used to answer).

4.3.5.2 Laboratory and Field Testing. The choice of a setting for conducting evaluation studies involves several considerations. Many, but not all, formative evaluation questions can be addressed more easily and efficiently in the controlled laboratory setting or in single or small student group tests (Pipe, 1965). Some summative questions can also be answered with laboratory testing, but frequently the need for realism and/or the lack of adequate (representative) test subjects and equipment except at a field site require field testing. Another practical consideration is the need to obtain data which will convince decision makers about the effectiveness of a product. Field tests carry more weight with real-world decision makers, even though such tests may be subject to greater sources of bias.

In designing an evaluation plan it is important to achieve an appropriate balance of field testing and laboratory testing. Laboratory evaluations offer the advantages of strict experimental control and convenience, but are subject to criticism because of the artificiality they introduce (Chapanis, 1967). Field tests offer the advantage of greater realism, but are harder to control, often more expensive than laboratory studies to conduct, and usually much more inconvenient than laboratory tests. A judicious mixture of laboratory and field testing is recommended. Field testing should be employed as necessary to determine or verify system design characteristics and to validate system effectiveness. Laboratory testing should be used for the bulk of formative testing and for system validation to the extent that it is logically defensible to do so.

4.3.5.3 Maintaining Objectivity. An important feature of any evaluation effort is the objectivity of assessment and the consequent credibility the study can generate in those who utilize it. To a certain extent, technically adequate evaluation and purely objective evaluation are incompatible. The individuals most qualified to conduct an evaluation on the basis of their knowledge and competence regarding the object or system to be evaluated generally have the strongest biases and vested interests regarding the outcome of the study. The problem of obtaining evaluation data that is both valid and objective represents a dilemma, therefore, for the evaluation designer. To some extent a trade off of these factors is needed to obtain reasonably unbiased, but still meaningful and representative, evaluative data. This problem can be addressed by a variety of approaches, including:

(1) Objective review of evaluation plans, procedures and data by a panel of independent consultants.
(2) Functional separation of responsibilities for evaluation activities from those for system design, development, and implementation.
(3) Use of double-blind experimental procedures whenever possible in individual evaluation studies.

All of the above procedures can be utilized in conducting an evaluation in which the system or product developer is also responsible for the evaluation. While these techniques are not foolproof, they should be carefully considered by evaluators in order to ensure that maximum objectivity and credibility are associated with evaluative findings.

4.3.5.4 Student and Instructor Needs. A common problem in evaluations of instructional products is overemphasis on the needs and activities of the student users and underemphasis or even neglect of the needs and activities of such secondary users as instructors, audiovisual and equipment personnel, and management personnel. Clearly student achievement, attitudes, and costs are important variables to consider in determining the impact of an educational innovation. At the same time the ultimate success of the product may depend as much on its acceptance by instructional, support and management personnel as on the demonstration of student achievement.

Questions of interest to secondary users include the robustness of equipment/methods in the field setting; the skills, costs, and time required to implement and maintain the system; the flexibility or loss of flexibility which accompanies introduction of the innovation into an ongoing curriculum or course; and the compatibility of the innovation pedagogically and philosophically with the ongoing instructional framework. The evaluation design must be sufficiently comprehensive and sensitive to allow assessment of these factors if they could play a significant role in the long-term success of the product under test.

It may be obvious that a similar concern applies to the design of field tests. Both the needs of the primary target audience (students) and the secondary target audience (instructors, support personnel and management) must be considered in designing and implementing field evaluations. All personnel should be properly briefed, provisions should be made for maintenance and logistical problems, efforts should be made to minimize disruption to other on-site activities, and the relevance of the test to ongoing activities and philosophies should be appropriately addressed.

4.4 Evaluating Decision Training

Goodman, Fischhoff, Lichtenstein, and Slovic (1976) describe three general approaches to evaluating decision training: outcome-oriented, process-oriented, and problem-oriented approaches.

The outcome-oriented approach is the traditional method for training evaluation and typically uses before-after comparisions of performance to demonstrate the effects of training. This approach is relatively straightforward if an external outcome measure exists. In the real world, however, with probabilistic decision situations, good decisions may result in bad outcomes, and vice versa. Without a large sample of behavior available for evaluation, then, evaluation of training solely on the basis of decision outcome can be misleading.

The problem-oriented approach focuses on the identification and correction of recurrent problems in decision performance, such as correction of biases in judgment. The process-oriented approach involves evaluating training in terms of the degree to which decision makers learn to follow procedural rules which are assumed to represent good decision making. Both of these approaches suffer from the problem that decision outcome is not considered.

Nickerson and Feehrer (1975) see the question of evaluating decision making and decision training in terms of effectiveness and logical consistency. Effectiveness is defined in terms of decision outcome. An effective decision is one that leads to the result desired by the decision maker. A logically sound decision is one in which the decision maker's choice is consistent with available information and the decision maker's values and goals. They indicate a clear preference for evaluating decision making from the standpoint of logical consistency:

> *Decision-making behavior should be evaluated in terms of its logical defensibility and not in terms of its effectiveness, inasmuch as effectiveness is found to be determined in part by factors beyond a decision maker's control, and usually beyond his knowledge as well. It often appears not to work this way in practice, however. Evaluation of decisions in terms of their outcomes seems to be the rule, for example, in the world of finance and business. Investment counselors are hired and fired on the basis of the consequences of their portfolio recommendations, and corporate managements are frequently juggled as a result of unsatisfactory profit and loss statements. Although the cliche "it's the results that count" has particularly strong intuitive appeal in this context, decision outcome is no more justified as the basis for evaluation of decision making in the financial world than in any other.* (p. 161)

At the same time, they recognize the realities of the operational environment which may be responsible for the traditional emphasis on assessing decisions on the basis of outcome:

It is probably safe to assume that <u>most</u> people in decision-making positions are more likely to be rewarded, or censured, as the case may be, on the basis of the effectiveness of their decisions than on that of their logical quality. This is due in part perhaps to the fact that society is far more interested in the results produced by its decision makers than in the reasons for which decisions were made. It is undoubtedly also true, however, that it is easier to determine the outcome of a decision than to determine whether the decision was logically justified at the time that it was taken. (p. 162)

Sidorsky and Simoneau (1970) present five criteria for evaluating decision training which deal with problems frequently encountered by decision makers. These are:

(1) Stereotypy, i.e., the tendency of a decision maker to respond in a manner that is unnecessarily correlated with some other factor(s) in the tactical situation. The response is thus rendered predictable.
(2) Perseveration, i.e., the tendency to persist with a particular response or interpretation after the accumulated data make a different response more reasonable.
(3) Timeliness, i.e., the extent to which the decision maker achieves a proper balance between the amount of time available and the amount of time taken to reach a decision.
(4) Completeness, i.e., the degree to which a decision maker avails himself of all relevant information.
(5) Series Consistency, i.e., the extent to which a decision maker responds consistently in a series of sequentially dependent or interrelated actions. (p. 5)

These criteria were used in a series of ASW training studies to evaluate trainee performance on particular problems and could serve as the basis for useful, specific feedback.

May, Crooks, and Freedy (1978) describe a study using computer-assisted instruction for training electronic troubleshooting. Four measures of troubleshooting performance were utilized which compared student decision performance with ideal performance:

(1) Relative competence--expected utilities of the trainee as compared to an expert.
(2) Relative consistency--stability of trainee behavior.
(3) Relative information gain--amount of information obtained by the trainee as contrasted to the total amount available.
(4) Relative information gain of considered alternative--amount of information considered for a decision in comparision to the amount potentially available for consideration.

This process-oriented approach to evaluation was used to generate feedback to trainees during and after specific circuit problems and was also used to select on an adaptive basis subsequent training problems which maximized experience in areas of student weakness.

Nickerson and Feehrer (1975) suggest a somewhat similar process-oriented approach to decision evaluation which is based on an eight-stage process model for decision making:

> *The adequacy of the information-gathering process; the sensitivity of data evaluation; the appropriateness of the structure that is given to a decision problem; the facility with which plausible hypotheses are generated; the optimality of hypothesis evaluation; the sufficiency with which preferences are specified; the completeness of the set of decision alternative that is considered; the timeliness of action selection and its consistency with the decision maker's preferences, objectives and information on hand.* (p. 163)

They point out that it is difficult to determine by post-hoc analysis whether a decision is logically sound. Once a decision is made <u>and</u> the results are observed, the decision maker can offer plausible reasons for the choices made. The validity of hindsight, which is subtly and perhaps unconsciously affected by outcomes, is questionable.

The question of evaluating decision processes on the basis of verbal reports has been considered extensively by Elstein, Shulman and Sprafka (1978). They combined objective measures, introspection, and judgments of experts to develop a comprehensive and reliable set of process measures for medical problem solving. Their approach appears to answer the concerns raised by Nickerson and Feehrer and offers guidance to others wishing to assess decision processes.

Training is generally conducted in an artificial setting--the classroom, laboratory, or simulator. Use of simulated real-world environments raises the question of the representativeness of the training environment and of the skills acquired. Sidorsky and Houseman (1966, p. 1) indicate that tactical training situations should not overemphasis favorable decision problems:

> *Training programs intended to develop general decision-making skills and characteristics related to combat situations should provide considerable practice in making tactical decisions in situations wherein the decision maker is at a disadvantage relative to the enemy. Uncertain, ineffective, and unrealistic behavior was found to occur in such situations in two separate experimental studies.*

In a similar vein, Bainbridge (1974) describes the effects of stress (personal danger) on subjects' ability to implement decisions. He reports that subjects were able to choose the appropriate action in simulations of the operation of an electric arc furnance, but "froze" when called on to execute the decision. Janis and Mann (1977) also discuss cases of decision avoidance which are related to conflict and stress. Given that many real-world decisions involve some element of risk, it seems important to train and evaluate decision behavior under representative circumstances of risk.

The ultimate test of any decision training program is its long-term effect on performance. Ultimately training should result in improved decisions and decision outcomes. As with most areas of training, long-term studies of the impact of decision training are lacking. This area represents an important, but neglected research topic.

Kanarick (1969) is representative of those who are concerned with decision making as a general skill. Clearly generalizability is an important area of evaluation, but one which is notably absent from studies of decision training. At a minimum, the evaluation of generalizability would require pre- and post-training assessments in tasks and contexts different from those in which training occurs. In addition, the development of generalized skills should be assessed in the real-world environment as a part of impact studies of long-term transfer effects.

In summary, the evaluation of decision making can involve both the effectiveness of decision outcomes and the procedural aspects of decision behavior. While a number of authors stress the importance of assessing decision process quality, evidence exists that trainees can learn to improve decision quality solely on the basis of outcome (e.g., Elstein, Shulman, and Sprafka, 1978) and that detailed feedback or coaching is not always beneficial. While it appears that the evaluation of decision processes can provide useful feedback to guide selection of subsequent training, the question of how best to use decision process evaluation data for instructional feedback is not settled. Clearly, decision training must be evaluated in terms of its impact on real-world performance. Impact studies of decision training could include assessments of the generality of decision skills learned.

4.5 Evaluating Aircrew Emergency Decision Skills

Although Prophet (1976) points out that little is known about training or evaluating such advanced aicrew skills as decision making, this topic has received increasing attention in recent years. Thorpe, Martin, Edwards, and Eddowes (1976) provide a paradigm for situational emergency training (SET) in which a library of emergency scenarios is used to provide trainees with exposure to a series of graduated difficulty exercises which involve a variety of situational factors. Evaluation is a central factor in SET instruction with student-instructor dialog used both to present situational detail and to clarify and correct student responses. SET can involve training in advanced simulators in addition to mock-ups or simple procedural trainers, with the introduction of additional realism and stress, and with the potential for simulating adverse outcomes.

Jensen (1978) discusses the evaluation of aircrew decision skills under the heading of judgment. Jenson (1978) defines judgment in terms of discriminative judgments and response selection tendencies in a manner similar to Sidorsky and Simoneau (1979). Three areas of judgment evaluation for aircrew members are described--pretraining evaluation, training evaluation, and transfer evaluation. Pretraining evaluation has to do with the screening and selection of those individuals with the highest potential for successful decision making. As Jensen (1978) indicates, little has been done in this area although psychometric assessments of risk-taking behavior may hold some promise for this area.

Training evaluation is treated by Jensen in terms of achievement of behavioral objectives, which leads back to the questions of how to define judgment in operational and emergency settings and how to train such skills. Both decision process and outcome measures could be considered relevant here. Transfer evaluation is conceived of as an extension of training evaluation and the results are viewed as being used to modify student selection criteria as well as training program design.

Murray and Brecke (1979) review proposals to implement aircrew decision training. They conclude that three principles can be used to improve current aircrew decision training:

> *(1) Reducing the domain of tasks which actually represent judgment performance.*
> *(2) Training this judgment performance using control of learning situations and academic presentation, and;*
> *(3) Considering (and, therefore, reducing) the effects of stress. (p. 461)*

Presumably stress induced in the training situation will be representative of that experienced under actual operational and emergency situations. Zavalova and Ponomareuko (1971) feel that the introduction of stress does not require high fidelity simulations, but can be approximated through a "collison of ideas" in which contradictory information is presented to the pilot. Their study of pilot responses to a malfunctioning autopilot appears to support this conclusion.

Zavalova and Ponomareuko (1971) propose the artificial creation of stress through the use of conflicting information as a means to test the reactions of aircrew members and screen out unsuited personnel. They do not report concrete evidence to support the validity of this approach, however.

Clearly, the evaluation of aircrew emergency decision training and skills has received insufficent attention. As Murray and Brecke (1979) point out, it is necessary to define precisely what constitutes aircrew emergency decision skills. Once this formidable task is completed, meta-evaluation, evaluation theory, and work on the evaluation of decision training and decision making in other fields can be profitably applied to aircrew emergency decision training evaluation.

5. SYNTHETIC LEARNING ENVIRONMENTS TO IMPROVE DECISION MAKING IN EMERGENCY SITUATIONS

5.1 Overview

Decisions taken by military pilots in situational emergencies are characterized as being unique and as providing little time for deliberation. The general concensus is that the standard method of decision making in these situations is to use intuition (Slovic, 1976), because application of the quantitative techniques of decision analysis is too time intensive. While it is not at all certain that increasing pilot's awareness of common decision making fallacies such as the base rate fallacy or the representativeness heuristic (see Slovic, 1976) will reduce the use of such erroneous decision making strategies or that nonspecific training in this area will generalize to the specific decision making that occurs in emergency situations, experience with selected exercises with synthetic learning environments should help pilots develop intuitive cognitive strategies to assist them in making better decisions in emergency situations. Such exercises should be designed to teach pilots the skills required to deal with unforeseen contingencies and should consist of conservative decision making procedures which allow for fast corrective action to recover from mistakes.

A prototype instructional approach which represents one method to provide pilots with experience in emergency decision making is described in detail in this chapter. Salient features of the instructional approach taken here are: (a) the use of inductive learning methods rather than deductive methods, and (b) the use of spiral learning (Pipe, 1966). Inductive learning leads students to develop appropriate cognitive strategies by having them perform specific exercises rather than teaching them abstract principles. Essentially, it is learning by doing. In spiral learning, the student first receives an overview of the whole concepts to be learned, and then is taken in a series of passes through the material. In this application, each pass requires the students to deal with increasingly complex synthetic learning environments. Furthermore, the pilot learns the cognitive strategies in their logical sequence--e.g., first to recognize a situational emergency; second, to identify the nature of the malfunction; third, to generate candidate decision alternatives; and, fourth, to select the best alternative.

The prototype exercise presented here is designed to use as input the cue patterns or emergency situation scenarios described in the previous chapters of this report. The approach presupposes that pilots are already familiar with preplanned standard procedures that can be applied to elementary emergency situations (e.g., boldface procedures). For the purpose of developing exercises to enhance cognitive decision making skills in situational emergencies, the decision making process is decomposed into four major stages. These are: (a) problem recognition, (b) malfunction identification, (c) alternative generation, and (d) alternative selection. More complete descriptions of the decision making process have been proposed in the literature (e.g., Nickerson & Feehrer, 1975; Schrenk, 1969); however, those descriptions were designed to describe the process of quantitative decision analysis, and are not readily applicable to the case of intuitive decision making.

The prototype learning exercise developed in the following sections is based on clearly defined performance objectives, i.e., the outcomes of the learning process are operationally defined such that they can be observed. Precision in the definition objectives meets two needs: (a) to communicate the purpose of the instruction and (b) to evaluate the efficacy of the instruction (Gagne & Briggs, 1979). Accordingly, in the following sections, the behav[illegible] objectives associated with each stage of emergency decision making process are first presented. The objectives are followed up by a prototype exercise which utilizes and builds on the behavioral objectives presented earlier.

5.2 Behavioral Objectives

The behavioral objectives which reflect on ability to deal with emergency situations are organized to correspond to the following four major goals: (a) recognize a situational emergency, (b) identify the nature of the malfunction, (c) generate candidate decision alternatives, and (d) select the best decision alternative.

5.2.1 Objective 1: Recognize a Situational Emergency. To recognize a situational emergency pilots must attend to a series of situational cues which are indicative of the emergency. These are best given in an aircraft simulator rather than by means of a paper and pencil exercise. Accordingly, such objectives are listed here for reasons of completeness, but are not included in the prototypical learning environment which are based on paper and pencil methods and which are given in the following section. Objectives based on situational cues are identified by the word "simulator."

Behavioral Objectives

(1) Attend to all situational cues (simulator).

- (a) Attend to aircraft factors (simulator).
- (b) Attend to environmental factors, e.g., weather (simulator).

(2) Identify cues that indicate a possible malfunction.

(3) Assess the apparent danger level of the situation.

(a) Judge potential risk of loss of life (own or other) or loss of aircraft.

(b) Judge time available to respond.

(c) Combine time and risk judgments to form overall situational evaluation.

(d) Choose one of the following courses of action: continue inventory, perform immediate action, or perform problem solving activity.

5.2.2 Objective 2: Identify Nature of Malfunction.

(1) Generate list of possible malfunctions in decreasing order or seriousness given cues.

(2) Identify additional information needed to confirm/disconfirm most serious and/or likely hypothesis.

(3) Obtain needed information.

(4) Confirm/disconfirm probable existence of malfunction.

(5) Repeat steps (2) to (4) for each serious and/or likely hypothesis.

(6) Determine whether all cues are explained by hypothesis.

(7) Repeat steps (1) through (6) for unexplained cues.

(8) Choose one of the following courses of action: continue monitoring situation, perform immediate action, or perform problem solving activity.

5.2.3 Objective 3: Generate Candidate Decision Alternatives.

(1) Review, rank-order, and select most important decision attributes.

(a) Develop candidate action(s) for given malfunction(s) and environment for the most important decision attributes (e.g., pilot life and survival, airplane and equipment cost, mission, peer factors, situational factors).

(b) Predict likely outcome(s).

(c) Add other acceptable actions to the list of candidate actions.

(d) Accept actions that do not have strong negative outcomes.

5.2.4 Objective 4: Select Best Decision Alternative.

(1) Determine most important decision attribute.

(2) Select decision alternative(s) having highest utility on this attribute.

(a) Estimate utility for each alternative for the given attribute.

(b) Rank decision alternatives on the basis of the most important attribute and select the highest ranked alternative(s).

(3) If two or more alternatives are given equal values in step 2b, repeat procedure given in step 2 for the next most important attribute, etc.

(4) Execute the best alternative.

(5) Continue to monitor the situation.

5.3 Prototype Learning Exercise

5.3.1 Part I.: Recognize a Situational Emergency. You are returning from a midday training mission in an F-4 aircraft at 2500 feet on a cloudy day. You are over land, 200 miles from the nearest landing area. You experience a bird strike and you notice the following:

(1) Fuel fumes in the cockpit.
(2) Master caution light goes on.
(3) Abnormal decrease in fuel.
(4) Low engine rpm in the left engine.
(5) Loss of thrust in the left engine.
(6) Lack of response in the left engine.
(7) Muffled banging sound.
(8) Autopilot disengage light goes on.

Which, if any, of these events have immediate importance for flight safety? (Please check or circle one in the list above.)

If you check (1) fuel fumes in the cockpit, (3) abnormal decrease in fuel quantity, (4) low engine rpm in the left engine, (5) loss of thrust in the left engine, (6) lack of response in the left engine, and (7) muffled banging sound, you are in agreement with expert flight instructors. The illumination of the master caution light and the autopilot disengage light indicate that you have a malfunction, but do not indicate the nature of the malfunction.

Given these events, how much risk do you rate in this situation in terms of loss of aircraft or personal injury? (Check two.)

- ☐ Very little risk of losing the aircraft and/or loss of life.
- ☐ Minor risk of losing the aircraft.
- ☐ Minor risk or losing your life.
- ☐ Moderate risk of losing the aircraft.
- ☐ Moderate risk of losing your life.
- ☐ High risk of losing the aircraft.
- ☐ High risk of losing your life.
- ☐ Certain or almost certain loss of aircraft and/or loss of life.

Expert flight instructors indicate that in this situation you face a high degree of risk of losing both the aircraft and of losing your life.

How much time do you feel you have to deal with this situation? (Check one)

- ☐ Very little - a few seconds.
- ☐ Little - a few minutes.
- ☐ Some - 5 to 10 minutes.
- ☐ A lot - 10 minutes or more.

Expert flight instructors indicate that you have <u>little</u> time (only a few minutes) to deal with this situation.

Given the risk involved in the situation and the amount of time you have to respond, what is your overall judgment of the danger involved in this situation? (Check one)

- ☐ Very dangerous.
- ☐ Dangerous.
- ☐ Somewhat dangerous.
- ☐ Not dangerous.

Expert flight instructors indicate that this situation is <u>dangerous</u> because of the <u>high risk</u> involved and because you have <u>little time</u> with which to deal with the situation.

Now that you have identified the situation to some degree, would you: (Check one.)

- ☐ Continue to monitor the situation.
- ☐ Try to find out what is wrong with the aircraft and think of the various alternative actions you might take.
- ☐ Take immediate action.

The expert flight instructors agree that the best course of action in this situation is to try to find out what is wrong with the aircraft and think of the various alternative actions you can take.

To "continue to monitor" the situation is, of course, not a reasonable response. There is an obvious danger which needs to be handled in a fairly short time period. To "take immediate action" is not reasonable either since you haven't really determined what action needs to be taken. The time you have could best be used in trying to identify the basic aircraft malfunction so that effective action can then be taken.

Summary

So far, we have covered the major steps that you should first take when an emergency occurs. These are:

(1) Identify events (malfunction symptoms) which indicate possible threat to your life and to the survival of the aircraft.

(2) Estimate the degree of risk and danger you face.

(3) Estimate the amount of time you have available to deal with the situation.

(4) Decide on one of the following courses of action:

- Continue to monitor the situation.
- Try to find out what is wrong with the aircraft and think of various alternative actions you might take to resolve the situation.
- Take immediate action.

5.3.2 Part Two: Identify Nature of Malfunction. Given the events described earlier (e.g., fuel fumes in the cockpit, master caution light goes on, abnormal decrease in fuel quantity, low left engine rpm,, loss of left engine thrust, lack of left engine response, muffled banging sound, and autopilot disengage light goes on) what do you suspect may be going wrong with the aircraft? List as many realistic possibilities as you can think of from most serious to least serious:

(a) ____________________
(b) ____________________
(c) ____________________
(d) ____________________
(e) ____________________

Likely malfunctions which could be associated with the cues given on the previous page are (from most serious to least serious):

(1) A fuselage fuel leak due to ruptured fuel cells or fuel lines.
(2) A compressor stall due to foreign object damage (FOD).
(3) Variable area inlet ramp failure.

What steps would you take to confirm or rule out the possibility of <u>a fuselage fuel leak</u>? List them below:

(1) ____________________
(2) ____________________
(3) ____________________
(4) ____________________

To determine whether you have a fuselage fuel leak, you could request your wingman to verify whether fuel is emitting from the fuselage. Assume that such a visual check has confirmed the fuel leak. You have also observed an abnormal decrease in fuel quantity and fuel fumes in the cockpit which both indicate a fuselage fuel leak.

How likely is it that you have a fuselage fuel leak. (Check one.)

- ☐ Very likely.
- ☐ Likely.
- ☐ So-so.
- ☐ Unlikely.
- ☐ Very unlikely.

Given the above factors, it is <u>very likely</u> that you have a fuselage fuel leak.

What steps would you take to confirm or rule out the possibility of <u>compressor stall</u>? List them here.

(1) ____________________
(2) ____________________
(3) ____________________
(4) ____________________

To determine whether you have a compressor stall, you could check to see if you have high exhaust gas temperature (EGT), an open nozzle, and lack of engine response to thrust. Assume that all of these indicators are positive for the left engine. Remember that you have also previously noted that the left engine has low rpm and that there exists a muffled banging sound.

How likely is it that you have a compressor stall. (Check one.)

- ☐ Very Likely.
- ☐ Likely.
- ☐ So-so.
- ☐ Unlikely.
- ☐ Very unlikely.

Given the above factors, it is very likely that you have a compressor stall.

What steps would you take to confirm or rule out the possibility of variable area inlet ramp failure? List them here.

(1) ______________________

(2) ______________________

(3) ______________________

(4) ______________________

To determine whether you have a variable area inlet ramp failure you could (1) observe the ramp position in the rear view mirror, (2) determine whether there exists significantly reduced fuel flow at power settings above 85% rpm, (3) determine if there is a high pitched howl at airspeeds above 300 knots, and (4) determine whether there is significantly reduced thrust (about 35%) at power settings above 90% rpm. Assume that all of these indicators are negative.

How likely is it that you have a variable area inlet ramp failure? (Check one.)

- ☐ Very likely.
- ☐ Likely.
- ☐ So-so.
- ☐ Unlikely.
- ☐ Very unlikely.

Given the above factors, it is very unlikely that you have a variable area inlet ramp failure. The best and clearest indicator of a variable area inlet ramp failure is visual inspection.

Up to this point, two malfunctions have been tentatively identified as the basis for the events noted:

(1) Fuselage fuel leak.
(2) Compressor stall.

Are there any events that cannot be explained by these two malfunctions. Review the list of eight events on page 1 and list those which are not explained by a fuselage fuel leak or a compressor stall.

(1) ______________________
(2) ______________________
(3) ______________________
(4) ______________________

If you answered "master caution light is on" and "autopilot disengage light is on" you are correct. However, these do not indicate a serious problem at this time and may be safely ignored.

At this point, would you (check one):

- ☐ Continue to monitor the situation.
- ☐ Develop a set of emergency actions that you might take and select the best one.
- ☐ Take immediate action.

According to expert flight instructors, the best strategy at this point is to develop a set of alternative emergency actions for dealing with this situation and to select the best one. This recommendation is made because you cannot ignore the implications of fuel loss and high EGT. The high EGT indicates that there is a serious possibility that the left engine might burn up. Also, you have identified two reasonably likely malfunctions which can be dealt with.

Summary

In the first part of this problem you decided whether you should (1) simply monitor the situation, (2) try to identify what was wrong with the aircraft and develop some alternative actions, or (3) take immediate action. In the second part of the problem your identified what the underlying malfunctions were, using these steps:

(1) Make up a list of possible malfunctions, given available information.

(2) Identify additional information needed to confirm/rule out the most serious and/or likely malfunction.

(3) Obtain the needed information.

(4) Confirm or rule out the probable existence of the malfunction.

(5) Repeat steps 2 to 4 for other serious and/or likely malfunctions.

(6) Determine whether all events are explained by the malfunctions identified.

(7) Seek additional information, if necessary, to explain any other serious but unexplained events.

(8) Decide whether to:

- Continue to monitor the situation.
- Continue solving the problem by developing and evaluating various alternative actions.
- Take immediate action.

5.3.3 Part Three: General Candidate Decision Alternatives. There are several factors to be considered when selecting from possible alternative actions in an aircraft emergency. Some alternatives may create a greater risk of death or serious injury, but may also increase the probability of mission success. For example, assume you are participating in a critical bombing mission. During the mission, your aircraft develops a malfunction where it consumes fuel very rapidly. You realize that you will have enough fuel to perform the mission, but not enough to return to home base.

Given the criticality of the mission, you might choose to perform the bombing run and later eject from the aircraft. This course of action may place you in greater risk than immediately returning to your base.

Critical factors that must be considered in choosing from the various alternative actions possible in a given emergency situation are:

(1) Personal safety - e.g., the probability that you will die or suffer serious injury.
(2) Cost - e.g., the cost of the aircraft and other equipment.
(3) Mission - e.g., the importance of the mission you are performing.
(4) Situation - e.g., the probability that your aircraft will crash into a school or residential area, etc.
(5) Personal Evaluation - e.g., whether your superior and fellow pilots will evaluate the emergency action you take as positive.

While these are the most important factors to consider, others may also be involved in particular situations.

In the current scenario, which are the important factors to be considered?

Review the five items on this page and list them here in decreasing order of importance.

(1) ____________________
(2) ____________________
(3) ____________________
(4) ____________________
(5) ____________________

Are there any other factors that should be considered? List them here.

(1) ____________________

(2) ____________________

Our expert flight instructors feel that because this is a training mission, the order in which factors should be ranked is:

(1) Personal safety.
(2) Cost.
(3) Evaluation.
(4) Situation.
(5) Mission.

In this situation (where you have identified a fuselage fuel tank leak and a compressor stall as likely malfunctions), what realistic actions can you take that will <u>minimize risks</u> to your <u>personal safety</u>? List them below:

(1) ____________________

(2) ____________________

(3) ____________________

(4) ____________________

For each possible action you have listed, what positive and negative outcomes do you think might occur if you took this course of action? List them here.

Possible Action	Positive Outcome	Negative Outcome
1	____________________	____________________
	____________________	____________________

Possible Action	Positive Outcome	Negative Outcome
2	______________________	______________________
	______________________	______________________
3	______________________	______________________
	______________________	______________________
4	______________________	______________________
	______________________	______________________

Expert flight instructors agree that a realistic action you can take in this situation to minimize risks to your personal safety is:

(1) To eject.

The positive and negative outcomes that might occur if you took these courses of action are:

Action	Positive Outcome	Negative Outcome
(1) Eject	Good chance of survival	Lose aircraft for certain

What realistic actions can you take that would reduce <u>costs</u> to the aircraft and other equipment? List them below.

(1) ______________________________
(2) ______________________________
(3) ______________________________
(4) ______________________________

For each possible action that you have listed, what positive and negative outcomes do you think might occur if you carried out the course of action?

Possible Action	Positive Outcome	Negative Outcome
1	____________________	____________________
	____________________	____________________
2	____________________	____________________
	____________________	____________________
3	____________________	____________________
	____________________	____________________
4	____________________	____________________
	____________________	____________________
5	____________________	____________________

Expert flight instructors agree that realistic actions you can take in this situation to reduce costs to the aircraft and other equipment are:

(1) Exercise standard emergency procedure for fuel leak.

(2) Exercise standard emergency procedure for compressor stall.

The positive and negative outcomes that might occur if you took these courses of action are:

	Action	Positive Outcome	Negative Outcome
(1)	Emergency procedure for fuel leak.	Good chance of saving aircraft. Reduce risk of losing life.	Chance of having a catastrophic fire. Reduction in aircraft range.
(2)	Emergency procedure for compressor stall.	Save left-hand engine from over-temperature Reduce risk of losing life.	Retain only right-hand engine.

In this scenario, the evaluation, situation, and mission factors are not as important and need not be considered in the decision making process.

At this point review all of the actions you might take and list them here:

(1) ____________________
(2) ____________________
(3) ____________________
(4) ____________________
(5) ____________________
(6) ____________________
(7) ____________________
(8) ____________________
(9) ____________________

Your list should look as follows:

(1) Eject.
(2) Execute emergency procedure for fuselage fuel leak.
(3) Execute emergency procedure for compressor stall.

Cross out on this list all actions which have one or more strongly negative outcomes.

You should have crossed out the following actions:

(1) Eject, because it would guarantee the loss of the aircraft.

Your final list of possible actions should look as follows:

(1) Execute emergency procedure for fuselage fuel leak.
(2) Execute emergency procedure for compressor stall.

Summary

In the first part of this problem you made a preliminary decision on how to approach the emergency. In the second part of the problem you used all available information to identify, as best you could, what the underlying malfunctions were. In this part of the problem you made up and evaluated several possible courses of action that could be adopted, using the following steps:

(1) Review the factors that are critical to choosing a course of action.

(2) Rank the factors (e.g., personal safety, cost, mission, situation, evaluation) from most critical to least critical.

(3) For each factor, make up a set of possible actions that can be taken.

(4) Consider the likely positive and negative outcomes that could occur if you took a given course of action.

(5) Reject any action that potentially has a strongly negative outcome.

(6) Make up a list of the remaining actions that do not have strong negative outcomes.

5.3.4 Part Four: Select Best Decision Alternative. You might recall that in the prevous section you considered the factors that were important in considering various alternative actions. In decreasing order of importance for this problem, these factors are:

(1) Personal safety.

(2) Cost.

(3) Evaluation.

(4) Situation.

(5) Mission.

Personal safety was ranked first because the emergency situation occurs in a training situation and your life is more important than saving the aircraft or any of the remaining three factors.

You may also recall that the two following alternative actions are under consideration:

(1) Execute emergency procedure for fuselage fuel leak.
(2) Execute emergency procedure for compressor stall.

Your task at this point is to select the best alternative action among the two under consideration. Among the two actions, which action or actions are best in regard to personal safety? List the number(s) of the action(s) here:

____

If you listed more than one option, go to the next page. Otherwise, go to the page that follows the next page.

You have identified several actions that can be taken that would be equally good in terms of personal safety. Your task is to now select one of these actions. The second most important factor in making your selection is cost. Which of these actions is best in terms of minimizing cost in this situation? List the number of the action here: ________.

If the actions are equal in terms of cost, choose the action that is best in terms of the third factor, namely, situation, and so on until one action remains. List the number of the selected actions here: ________.

If you chose action number one you are in agreement with expert flight instructors. This is the best action to select under these circumstances, given that personal safety is the most important factor in this situation. At this point you would normally carry out the selected action. However, you should continue to monitor the aircraft's performance and also the situation after taking this action to be sure that the malfunction you have identified accurately reflects the aircraft's real malfunction and that the action taken is the proper response to the situation.

In this scenario, you would also take action number two to insure that you did not lose the left-hand engine to overtemperature.

Summary

In the first step of this problem you made a preliminary decision on how to approach the emergency. In the second part of the problem you used all available information to identify, as best you could, what the underlying malfunctions were. In the third part of the problem you generated and evaluated several possible courses of actions and rejected the ones with strongly negative outcomes. In this part of the problem, you selected the best decision option using the following steps:

(1) Review the most important factor (personal safety in this case) in choosing an action.
(2) Select the action that is best in regard to this factor.
(3) If several possible actions are equal in terms of this factor, decide which of these actions is best in terms of the second most important factor (e.g., cost). If the possible actions are still equal, continue to each remaining factor until one action remains.
(4) Carry out the action selected.
(5) Determine whether the action taken has the desired outcome by continuing to monitor the situation.

5.4 Conclusion

This chapter describes a prototypical instructional approach to provide pilots with experience in emergency decision making. To be used in the field, this approach should be expanded in several directions. First, the exercises should expand to encompass a variety of flight phases and environmental factors--e.g., takeoff, landing, high altitude, low altitude, bad weather, etc. Second, different cue patterns should be used to provide pilots with exposure to emergency situations which vary in: (a) the severity of malfunctions, (b) the uniqueness of malfunctions, and (c) the difficulty of the cue pattern to be interpreted. Third, the problems should be designed to differentially emphasize one or more phases of the decision process--e.g., problem recognition, malfunction identification, alternative generation, etc. Finally the problems should be designed to model cascaded decision situations--i.e., situations where the current decision is dependent on a prior decision and where the current decision will affect later decisions. Generally, emergency situations require the ability to make a series of decisions rather than a one-shot decision that completely handles the emergency.

6. REFERENCES

Bainbridge, L. Analysis of Verbal Protocols from a Process Control Task. In E. Edwards and F.P. Lees (eds.) The Human Operator in Process Control. London: Taylor and Francis, 1974, 407-450.

Bartlett, F.C. Remembering. Cambridge, England: Cambridge University Press, 1932.

Campbell, D.T. and Stanley, J.C. Experimental and Quasi-Experimental Designs for Research on Teachings. In N.L. Gage (ed.) Handbook of Research on Teaching. Chicago: Rand McNally, 1963, 171-246.

Chaparris, A. Research Techniques in Human Engineering. Baltimore: John Hopkins, 1958.

Chaparris, A. The Relevance of Laboratory Studies to Practical Situations. Ergonomics, 1967, 10, 557-577.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1969.

Edwards, A.L. Experimental Design in Psychological Research. (Rev. ed.) New York: Holt, 1960.

Elstein, A.S., Shulman, L.S. and Sprafka, S.A. Medical Problem Solving. Cambridge, MA: Howard University, 1978.

Gagne R., Briggs, L. Principles of Instructional Design. (Second ed.) New York: Holt, Rinehart & Winston, 1979.

Goodman, B., Fischhoff, B., Lichtenstein, S. and Slovic, P. The Training of Decision Makers. Report of a conference held at the Oregon Research Institute, Eugene, Oregon, July 8-9, 1976.

Hebb, D.O. The Organization of Behavior. New York: Wiley, 1949.

Janis, I.L. and Mann, L. Decision Making. New York: Free Press, 1977.

Jensen, R.S. Pilot Judgment: Training and Evaluation. Proceedings of the 11th NTEC-Industry Conference, 1978.

Kanarick, A.F. The Learning, Retention, and Transfer of Decision Making. Paper presented at meetins on Learning, Retention and Transfer, sponsored by Honeywell Systems and Research Center and the Naval Training Device Center (Winter Park, FL), 1969.

Kerlinger, F.N. Foundations of Behavioral Research. New York: Holt, 1964.

Lashley, K. Cerebral Organization and Behavior. Proc. Assoc. Res. Nervous Mental Disease, 1958, 36, 1-18.

May, D.M., Crooks, W.H. and Freedy, A. Application of Adaptive Decision Aiding Systems to Computer-Assisted Instruction: Experimental Studies. Technical Report PART-1023-76-1, Perceptronics, 1976.

Miller, G.A., Galanter, E. and Pribram, K.H. Plans and the Structure of Behavior. New York: Holt, Rinehart and Winston, Inc., 1960.

Moursund, J.P. Evaluation: An Introduction to Research Design. Monterey, CA.: Brooks/Cole, 1973.

Murray, S.A. and Brecke, F.H. Implementing Aircrew Judgment Training. Proceedings Interservice-Industry Training Equipment Conference, NAVTRADEVCEN Technical Report IH-316, 1979.

Nickerson, R. and Feehver, C. Decision Making and Training. Naval Training Equipment Center, Orlando, FL. Technical Report NAVTRAEQUIPCEN 73-C-0128-1, July 1975.

Parsons, H.M. Man-Machine System Experiments. Baltimore: Johns Hopkins, 1972.

Pipe, P. Practical Programming. New York: Holt, 1965.

Popham, W.J. Evaluation in Education. Berkeley, CA.: McCutchan, 1974.

Prophet, W.W. Long-Term Retention of Flying Skills: A Review of the Literature. Hum RRO Technical Report FR-ED(P)-76-35, October 1976.

Reynolds, J., Neuse, M., Olds, S. and Levine, M. Evaluation Handbook for Family Planning Programs. USDHEW, PHS, HSA, BCHS, Rockville, MD., 1978.

Schrenk, L.P. Aiding the Decision Maker--A Decision Process Model. Ergonomics, 1969, 12, 543-558.

Scriven, M.S. The Methodolgoy of Evaluation. AERA Monograph Series on Curriculum Evalution. Chicago: Rand McNally, 1967.

Scriven, M.S. An Introduction to Meta-Evaluation. Educational Product Report, 1969, 2, 36-38.

Scriven, M.S. Evaluation Perspectives and Procedures. In W.J. Popham (Ed.) Evaluation in Education. Berkeley, CA.: McCutchan, 1974, 1-94.

Sidorsky, R.C. and Houseman, J.F. Research on Generalized Skills Related to Tactical Decision Making, Naval Training Device Center, Port Washington, N.Y., Technical Report NAVTRADEVCEN 1329-2, December 1966.

Sidorsky, R.C. and Simoneau, G.R. An Experimental Evaluation of Tactrain. Naval Training Device Center, Orlando, FL. Technical Report NAVTRADEVCEN 1329-4, June 1970.

Slovic, P. Towards Understanding and Improving Decisions. In Science, Technology, and the Modern Navy, Thirtieth Anniversary, 1946-1976. Arlington, VA: Department of the Navy, Office of Naval Research, 1976.

Stufflebeam, D.L. Meta-Evaluation. Paper No. 3 in the Occasional Paper Series. College of Education, Western Michigan University, Kalamazoo, Michigan, 1974.

Tien, J.M. Toward a Systematic Approach to Program Evaluation Design. IEEE Transactions on Systems, Man and Cybernetics, 1979, 9, 494-515.

Thorpe, J.A., Martin, E.L., Edwards, B.J., and Eddowes, E.E. Situational Emergency Training. Air Force Human Resources Laboratory (Brooks AFB) Interim Report AFHRL-TR-76-47(I), 1976.

Weiss, C.H. Evaluation Research. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Zaralova, N.D. and Ponomarenko, V. A. Characteristics of Flier's Behavior Under Complicated Flight Conditions, Vosprosy Psikhologii, 1970, 5, 110-121.

DAT
FILM